Vol. 52 AUGUST 1961 No. 4

Journal of Educational Psychology

CONTENTS

An Analysis of the Factorial Dimensions of the Achievement Motivation Construct 179
JAMES V. MITCHELL, JR.

Categorization and Related Verbalizations in Deaf and Hearing Adolescents 188
SOLIS L. KATES, WILLIAM W. KATES, JAMES MICHAEL, AND TERRENCE M. WALSH

Cross-Validation Groups, Extreme Groups, and the Prediction of Academic Achievement 195
LEONARD J. KAZMIER

A Note on Relationships between Student and Parent Minnesota Teacher Attitude Inventory Scores 199
ROBERT E. CAMPBELL AND JOHN E. HORROCKS

Test Anxiety and the Intellectual Performance of College Students 201
IRWIN G. SARASON

Effects of Memory, Evaluative, and Creative Reading Sets on Test Performance 207
E. PAUL TORRANCE AND JUDSON A. HARMON

The Effects of Visual Discrimination Pretraining with Word and Letter Stimuli on Learning to Read a Word List in Kindergarten Children 215
SIEGMAR MUEHL

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

JOHN B. CARROLL
Harvard University
LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College for Teachers
DOROTHEA MCCARTHY
Fordham University
CARSON MCGUIRE
University of Texas
PAUL H. MUSSEN
University of California, Berkeley
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University

The *Journal of Educational Psychology* publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $10.00 (Foreign $10.50). Single copies, $2.00.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-Class postage paid at Baltimore, Md.

Copyright, 1961 by the American Psychological Association, Inc.

JOURNAL OF

EDUCATIONAL PSYCHOLOGY

Volume 52 August 1961 Number 4

AN ANALYSIS OF THE FACTORIAL DIMENSIONS OF THE ACHIEVEMENT MOTIVATION CONSTRUCT

JAMES V. MITCHELL, JR.

University of Texas

The construct of achievement motivation has been the object of considerable research and discussion in recent years (Atkinson, 1958; McClelland, Atkinson, Clark, & Lowell, 1953), and it is only too apparent that there are wide differences of opinion about the adequacy of current efforts to measure that intriguing variable (Atkinson, 1960; Blake & Mouton, 1959). Much of that effort, of course, has been focused on the assessment of n Achievement by means of analyses of TAT protocols, a technique developed by McClelland et al. (1953) and now regarded as *sui generis* by its devotees. One adherent of this technique has even relegated all other methods of assessing achievement motivation to the limbo of "ad hoc" devices, and he admonishes all users of such devices not to assume that the theoretical conceptions and factual information resulting from projective analyses of n Achievement are equally pertinent to studies involving the "ad hoc" assessment devices (Atkinson, 1960, pp. 275–276). Others not quite so identified with the McClelland approach seem equally insistent that n Achievement scores from different instruments exhibit significant relationships to one another and to the common behavioral data they are supposed to measure. With discouraging regularity, however, investigators have reported the absence of relationship between different measures of achievement motivation (Bendig, 1957; Himelstein, Eschenbach, & Carp, 1958; Melikian, 1958), and attempts to predict actual behavioral criteria have often resulted in essentially negative findings, confusion, or vast conceptual reinterpretation (Birney, 1958; Lazarus, Backer, Broverman, & Mayer, 1957; Vogel, Baker, & Lazarus, 1958). It is readily apparent that there is a rather pronounced lack of consensus about both the conceptual and methodological aspects of the achievement motivation construct, and there is an urgent need for additional research that will help to provide structure for its currently amorphous state of existence.

The present investigation came about as a result of a larger study of the cultural and situational determinants of achievement motivation for a group of college students. For this larger study the measurement of achievement motivation was obviously a critical operation, and since the research literature provided ample evidence of the difficulties involved in this kind of assessment, it seemed imperative to select as many different measures of achievement motivation as would be necessary to assure a thorough representation of that domain and then to make a careful analysis of their correlates and factorial dimensions. It was early recognized that such

a procedure should result in findings that would have general theoretical significance and applicability, because they would help to shed light on whether achievement motivation was a unitary construct with invariable meaning, as many seem to regard it, or whether it referred in fact to several psychological dimensions, each with different implications and different empirical correlates. It is this task that defines the purpose of the present investigation.

Procedure

The following instruments were employed in the analysis to represent the domain of possible achievement related attitudes.

1. McClelland Test of Achievement Motivation. This instrument was administered and scored in accordance with the instructions given by McClelland in his book *The Achievement Motive* (McClelland et al., 1953). In scoring a representative sample of protocols provided in that book, the present investigator established that his scores correlated .92 with those reported by the test author.

2. A sentence completion test of achievement motivation. This test was developed by the investigator as a simplified projective measure of achievement motivation, with the thought that such a measure might well have more validity than an instrument with the scoring complexity of the McClelland test. The instrument consisted of 46 items, some developed by the author and some taken from experimental tests developed by other investigators. The items selected represented various degrees of achievement orientation, from strongly achievement oriented to complete lack of achievement orientation. The instrument was scored by applying to each item the very first criterion of the McClelland system for scoring the TAT protocols, viz., the presence or absence of an "achievement goal." According to McClelland, the existence of an achievement goal is indicated by any statement which implies "success in competition with some standard of excellence" (McClelland et al., 1953, p. 110). If such were expressed or implied in the completion of any item, that item was scored +1. If the item completion involved no achievement imagery of this type, it was scored 0. If the item completion expressed or implied an evident lack of success or interest in competing with some standard of excellence, it was scored −1. The score on the test was simply the algebraic sum of the item scores.

3. An adjective checklist containing achievement related items. A few achievement related adjectives were added to those similarly oriented and already present in the Bills' Index of Adjustment and Values (Bills, Vance, & McLean, 1951), and for each of these adjectives the subject rated himself on a 5-point scale. The subject's achievement motivation score on this instrument was the total of the numerical ratings for the contributing adjectives. Corrected split-half reliability for this score was .84. Remaining items not contributing to this score served to mask the real purpose for which the instrument was being administered. Noncontributing items also served this same function for the two following instruments.

4. An inventory-type test of achievement motivation. This inventory contained items of the True-False variety, many of which were borrowed from the MMPI. Other items related to achievement motivation were added by the investigator. The corrected split-half reliability of this instrument was .75.

5. A multiple-choice questionnaire including achievement motivation items. The alternatives for each item in the questionnaire were serially ordered so as to define choices along a continuum representing greater or lesser degrees of the variable in question. The number of an alternative chosen thus defined a position on the continuum and the individual's score for that particular item. The corrected split-half reliability of this instrument was .83.

6. Taylor Manifest Anxiety scale (1953). This scale was included in the battery to clarify whether the imputed drive properties of manifest anxiety, as hypothesized by the disciples of Hull, would receive *de facto* verification from the factor analysis of these achievement motivation variables.

7. Level of aspiration measures. All subjects were members of the investigator's sections in elementary educational psychology. Before the first and last examinations of the semester all subjects were asked to record their levels of aspiration by answering the question: "What grade can you reasonably hope to attain on this examination?" These aspiration estimates were originally received in letter grade form and then later converted to a 10-point numerical scale which recognized + and − distinctions, e.g.,

C+, B−, etc. Actual grade received on the examination was also converted to the same 10-point numerical system. For each subject the numerical value of the actual grade received was subtracted from the numerical value of his level of aspiration estimate. This was done for both the first and last examinations, resulting in two indices which represented the extent to which the subject actually attained or approached his level of aspiration.

8. A measure of discrepancy between achievement and potentiality. The numerical grade averages for all subjects for the semester in which the study was conducted were computed, normalized, and converted into standard score form. The ACE scores for all subjects were similarly normalized, and converted into standard score form. A discrepancy score was computed for each subject by subtracting his standard score for the ACE from his standard score for his semester grade point average (GPA − ACE).

All the afore-mentioned data were secured from 131 female college students who were in a teacher training curriculum. The analysis was divided into two parts: (*a*) The scores from the various instruments were correlated with one another and with grade point average criteria, and (*b*) as a result of the insights gained from the first analysis, all of the individual items comprising the various instruments, some 29 variables in all, were then intercorrelated, subjected to a centroid factor analysis as described by Thurstone (1947), and subsequently rotated to orthogonal simple structure by means of Kaiser's varimax technique (1958).

Results and Discussion

Table 1 shows the intercorrelations between measures of achievement motivation, ACE, and GPA. The GPA−ACE discrepancy score (last column) here functions as the criterion variable for achievement motivation. Immediately apparent are the differences in correlations between this criterion and the various measures of achievement motivation. The simple adjective checklist was the best predictor of the criterion. The achievement motivation inventory was significantly related to the criterion, but, unlike the adjective checklist, had a significant negative correlation with ACE as well. The level of aspiration measures were related significantly to the criterion, but this may be partly artifactual in nature, resulting from the presence of a common component, grades, in the make-up of predictor and criterion. Manifest anxiety was quite unrelated to the criterion. Neither of the projective instruments (sentence completion and TAT n Achievement) was related to the criterion. Interestingly enough, however, the sentence completion test did have a significant and moderately high negative correlation with ACE and a lower but still significant negative correlation with GPA. These correlations, along with the negative correlation reported earlier for the achievement motivation inventory and ACE, re-emphasize the necessity for cautious empirical inquiry into the real dimensions and correlates of complex measures of complex constructs like achievement motivation.

The intercorrelations between the various measures of achievement motivation included in Table 1 show some interesting and often perplexing trends. The adjective checklist is significantly correlated with the achievement motivation inventory but not with the sentence completion test. The sentence completion test is, however, significantly correlated with the achievement motivation inventory. The pattern indicates some overlap between instruments that presumably measure the same variable, but hardly inspire confidence in the unidimensionality of the achievement motivation construct. The level of aspiration measures, besides being related to the criterion, are also related significantly to one or two other measures of achievement motivation. The TAT and manifest anxiety variables show little relationship to other measures.

TABLE 1

INTERCORRELATIONS BETWEEN MEASURES OF ACHIEVEMENT MOTIVATION, ACE, AND GPA

Measures	Achievement Motivation Inventory	Sentence Completion	TAT n Achievement	Aspiration-Grade[a] (first test)	Aspiration-Grade[a] (last test)	Manifest Anxiety	ACE	GPA	GPA-ACE
Adjective Checklist	.40**	.13	−.07	.15	.02	−.15	−.02	.36**	.40**
Achievement Motivation Inventory		.35**	−.07	.07	−.20*	.08	−.23**	−.06	.18*
Sentence Completion			−.04	−.18*	−.06	.09	−.32**	−.17*	.09
TAT n Achievement				.13	.18*	.01	.11	.08	−.07
Aspiration-Grade[a] (first test)					.26**	.04	.20*	.48**	.27**
Aspiration-Grade[a] (last test)						.05	.19*	.48**	.27**
Manifest Anxiety							−.27**	−.11	.09
ACE								.44**	

[a] Direction of scoring reversed by subtracting each discrepancy from a constant; the higher scores represented the smaller discrepancies.

* Significant at .05 level.

** Significant at .01 level.

The whole pattern of interrelationships was one that suggested not only the multidimensionality of the putative achievement motivation construct, but also the probability that various measures of that construct might reflect quite different aspects of it and would therefore either be little correlated or selectively correlated with other measures and with the criterion. A factor analysis of all of the items of these various measures seemed admirably suited to shed light on such questions as these, and such an analysis was undertaken as the second step in the investigation.[1]

[1] The correlation matrix for the 29 individual achievement motivation items that were subjected to factor analysis has been deposited with the American Documentation Institute. Order Document No. 6739 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

Table 2 shows the rotated factor matrix which resulted when the *individual items* of the achievement motivation measures were intercorrelated and factor analyzed. The items of the multiple-choice inventory, not included in the earlier analysis, were included in the present one to broaden the basis for analysis and comparison. Factorization was terminated when the factors isolated accounted for approximately 100% of the estimated common factor variance. A factor was not interpreted, however, unless the product of its two highest loadings exceeded two standard errors of a zero-order correlation for an N of the magnitude used in this study (Fruchter, 1954, p. 79). Two criteria were used for the rotational procedure: the varimax criterion (Kaiser, 1958), and the requirement that the final rotational solution achieve optimum factorial simplicity for the GPA−ACE criterion. After the first criterion was met, however, there was only one small ad-

TABLE 2

ROTATED FACTOR MATRIX FOR 29 INDICES OF ACHIEVEMENT MOTIVATION

Measures	I	II	III	IV	V	VI	VII
1. GPA-ACE	74	01	00	05	−20	08	09
2. Aspiration-Grade Discrepancy[a] (first test)	30	−07	−32	30	−05	10	−14
3. Aspiration-Grade Discrepancy[a] (last test)	37	−04	−35	−11	−16	−06	−05
Adjective Checklist Items:							
4. Ambitious	25	09	33	36	−22	20	26
5. Industrious	28	13	42	43	−32	17	09
6. Hardworking	43	13	58	26	−19	−09	18
7. Studious	56	07	36	40	04	04	16
8. Competitive	11	34	08	45	11	−01	08
9. Motivated	44	15	28	60	−05	03	10
Achievement Motivation Inventory Items:							
10. I am more ambitious than the average person.	32	04	23	29	−04	31	28
11. When I'm faced with difficult tasks, I generally "give up" too easily.	−06	00	−05	−52	05	01	14
12. I exert considerable effort to get good grades.	37	12	57	03	06	28	09
13. I wish I were more successful as a student.	−18	−33	23	−20	23	06	−29
14. I try very hard to do well in anything I undertake.	29	06	18	05	12	54	12
15. I find that the standards I set for my own work are usually too high for me to attain.	01	−31	11	−15	55	17	−08
16. My parents exert considerable pressure on me to achieve good grades.	−27	01	07	10	49	04	−08
17. I have strong motivation to succeed in my work.	49	03	40	17	20	38	−17
18. I feel under constant pressure to do well in my school work.	04	02	16	−01	53	04	02
19. Manifest Anxiety	13	−36	−09	−11	30	−17	09
20. Sentence Completion	09	−03	33	−26	14	34	07
Multiple-Choice Questionnaire Items:							
21. Do you feel that thus far you have really lived up to your parents' expectations for you?	33	35	07	20	−41	01	−17
22. Are your parents satisfied with your grades in college?	59	27	04	27	−41	−11	−09
23. How hard do you work on your schoolwork in comparison with the typical student of your sex?	75	05	44	00	05	−02	−03
24. How important are *grades* to you?	34	−04	48	−03	06	15	−02
25. Do you feel that your parents are satisfied with you as a person?	17	65	13	−02	−20	01	−02
26. Are you satisfied with yourself as a person?	−01	69	−04	20	−01	02	04
27. Do you and your parents agree on "values"—what constitutes the really important things in life?	−03	48	24	−22	11	01	24
28. How would you rate your study habits in comparison with the average student of your sex?	70	15	18	13	−07	10	13
29. TAT n Achievement	−01	−02	−09	05	00	−06	−41

[a] Direction of scoring reversed by subtracting each discrepancy from a constant; the higher scores represented the smaller discrepancies.

ditional rotation that had to be made to meet the requirements of the second. The following factors were tentatively identified.

Factor I had its highest loadings for the GPA−ACE criterion and for two variables (23 and 28) which reflect the actual effort expended to do an effective academic job. Other variables have loadings to the extent that they also reflect an earnest application of effort to obtain tangible academic results. This factor might best be called Academic Motivation and Efficiency. It is interesting to note that the variables having the highest loadings on this criterion factor were objective self-rating indices, that the level of aspiration measures had moderate loadings on the factor, that the two projective measures (20 and 29) did not have loadings on this factor, and that neither manifest anxiety (19) nor self-satisfaction (26) had loadings on this important criterion factor.

Factor II had its two highest loadings for items relating to self-satisfaction and acceptance (25 and 26). It has negative loadings for manifest anxiety (19) and for the item expressing a desire to be more successful as a student (13). It seems appropriate to refer to this factor as Self-Satisfaction. It is interesting to note the lack of relationship between this Self-Satisfaction factor and the GPA−ACE criterion.

Factor III can best be interpreted in terms of its contrast with Factor I. It has no loading for the GPA−ACE criterion. It has negative loadings for the level of aspiration measures, indicative of its correlation with larger discrepancies between aspiration and achievement, as contrasted with the positive loadings of the first factor. It has a positive loading for the item expressing the desire for more success as a student, in contrast to the first factor's negative loading for the same item (13). Its loading of .04 contrasts with the first factor loading of .59 for the item expressing parental satisfaction with the subject's grades (22). Its much smaller loadings for Items 23 and 28, earlier interpreted as reflecting actual willingness to expend effort to achieve results, must be contrasted with the corresponding very high loadings for the first factor. For most of the remaining variables there seems to be a similar tendency for low loadings to be associated with items referring explicitly to effective application to academic work, while the items with high loadings are more vague and generalized, less behavior oriented, and more susceptible to a positive response set. All the evidence, and particularly that already noted for Items 1, 2, 3, and 13, suggests that Factor III should be interpreted as Wish-Fulfillment Motivation. The factor seems to represent strong desire but little application or fulfillment.

It is interesting to compare the third factor loadings of the sentence completion test with its first factor loadings. The resulting contrast strongly suggests that projective test users must avoid jumping to conclusions about what level of experience or motivation is being tapped by a projective test. The appreciable loading of the sentence completion test on the Wish-Fulfillment factor, its lack of loading on the first factor, and the earlier reported correlation of −.32 between sentence completion and ACE, all suggest that this projective may tend to identify the less able student who yearns for high achievement but who, because of discouragement or inertia, cannot really buckle down to achieve results. This hypothesis also receives partial confirmation from the original correlations.

Factor IV has a high positive loading for the adjective "motivated" (9) and a high negative loading for the in-

ventory item indicating the subject's tendency to "give up" with difficult tasks (11). Several lesser loadings represent similarly achievement oriented variables. What most distinguishes this factor, however, is its lack of loadings for all items related strictly to academic motivation alone. Items having no loadings on this factor include two items concerned with willingness to work hard on schoolwork (12 and 23), one concerned with the importance of grades to the subject (24), and one concerned with the effectiveness of the subject's study habits (28). Because of this specific exclusion of academic motivation from what seems to be a generally positive achievement orientation, it seems logical to designate this factor as Nonacademic Achievement Motivation. Appropriately enough, this factor has no loading for the GPA—ACE criterion.

Factor V is notable for its high loadings for items concerned with feeling under parental pressure to do good academic work (16 and 18), aiming at standards that are too high to attain (15), but still not living up to parental expectations or achieving grades satisfactory to them (21 and 22). This factor seems most clearly to define an External Pressure to Achieve.

Factor VI, the most difficult factor to interpret, has a high loading for an item referring to the subject's willingness to try hard in all he undertakes (14) and lesser but still appreciable loadings for items reflecting a generalized desire to achieve (10, 12, and 17). Yet it is quite unrelated to the GPA—ACE criterion. Unlike the third factor, it has no loading for the level of aspiration measures (2 and 3) or for the item reflecting a desire for more success as a student (13). Tentatively we will call this factor Imputed Generalized Motivation without Attendant Effort. It is interesting to note that the sentence completion test again loads on a factor that involves desire to achieve but little evidence of real or appropriately directed effort.

Factor VII should probably be interpreted as a residual or error factor of no importance. The product of its two highest loadings did not exceed the two standard error criterion. Its highest loading was for the TAT n Achievement test, whose appearance on this error factor is most probably due to low reliability. This same test had but one significant correlation in the entire battery, and this was low and perhaps attributable to chance.

An analysis of this type would seem to have several implications requiring the thoughtful consideration of researchers in this area. First, it is evident that the achievement motivation construct is *not* a unitary construct with invariable meaning, and that any attempt to regard it in this light is likely to lead to methodological confusion, conflicting results, and conceptual stagnation. In this study it was possible to isolate five or six relatively independent dimensions which represented different and psychologically significant aspects of the desire to achieve. It would therefore seem imperative for each research worker in this area to offer some empirical evidence of which of several possible varieties of achievement motivation he had chosen to employ in his investigation. To ignore the complexity of the achievement motivation construct is to render many results uninterpretable.

Secondly, it is apparent that tests and test items showing equivalent initial promise may turn out to differ widely in their ability to measure any given aspect of achievement motivation. Empirical surprises seem almost the rule rather than the exception. The TAT n Achievement test had the lowest communality of all 29 variables employed in the factor analysis, and

it loaded only on an error factor. Its one significant correlation in the original correlation matrix was a negative correlation with an item relating to ambition, and its correlation with the GPA−ACE criterion was −.07. On the other hand, two individual objective items (23 and 28) had very high loadings on the first factor and were correlated .53 and .52, respectively, with the GPA−ACE criterion. This is a notable result, not only because such correlations are even difficult to achieve with whole tests, but also because they were achieved with objective, multiple-choice, self-rating items that focus on concrete, easily observed behaviors that are a part of everyday experience.

Thirdly, tests and test items may differ widely in terms of which particular aspects of the desire to achieve complex they actually measure. For example, the sentence completion test had a .09 loading on the first factor, a correlation of .09 with the GPA−ACE criterion, and significant correlations of −.32 and −.17 for the ACE and GPA, respectively. There is certainly no evidence of its validity as a measure of achievement motivation here, and if GPA had been used as a criterion, results would be directly contrary to hypothesis. But the complete analysis revealed that this test was loaded on the Wish-Fulfillment factor, and its loading there gave credence to Lazarus' (Lazarus et al., 1957) contention that there is no one-to-one relationship between the expression of a need in fantasy and its actualization in behavior. Indeed, the fantasy may even serve as a substitute for overt striving. Analyses of fantasy, then, may not always be "depth" analyses; they may actually reveal relatively superficial aspects of behavior and personality.

Certain objective items also turned out to measure quite different aspects of the desire to achieve complex than had been originally anticipated. It had been anticipated, for instance, that Items 8, 11, and 18 would be highly related to the criterion. As it developed, they were not at all related to the criterion nor to the first factor; they were, however, loaded on other factors. Items 22 and 28, on the other hand, were highly related to the GPA−ACE criterion, having correlations of .50 and .52, respectively, but their patterns of loadings in the factor matrix were quite different. Here, too, cautious analysis is needed to ascertain just what aspect of the desire to achieve complex is actually being measured.

The importance of such analysis is particularly apparent in comparing the lack of correlation between some of the instruments represented in Table 1 with the relatively high loadings of the factor matrix and the relatively high interitem correlations of the original correlation matrix. A comparison of this kind furnishes compelling evidence that intertest correlations can be attenuated by items representing different factorial dimensions or by same-test items actually counteractive in their influence.

Fourthly, it is not only important to acknowledge that different tests and test items may represent different aspects of the desire to achieve complex; it is equally important to recognize that the *same* test or test item may represent several different aspects of the desire to achieve complex. The adjective "industrious" (9), for instance, has loadings on four different factors and could represent four distinctly different attitudes toward achievement. There are several other items (e.g., 9, 17, and 23) that are loaded on two or three factors. It is apparent that identical responses can have very different meanings for different subjects.

Finally, there were certain by-products of this investigation that are worthy of note despite their peripheral status. The pattern of loadings for manifest anxiety did not lend confirmation to the hypothesized drive attributes of that variable, but it must be acknowledged that this investigation was not designed as a critical test of that issue. It is also interesting to note that the item relating to self-satisfaction (25) was correlated .00 with the GPA—ACE criterion and .09 with GPA alone. The Self-Satisfaction factor had similarly insignificant and negligible loadings for items concerned with strictly scholastic motivation (7, 17, 23, and 24). Apparently the self-concepts of our college students are not nearly as influenced by matters relating to academic motivation and achievement as we are sometimes inclined to believe.

Summary

The purpose of this study was to determine whether "achievement motivation" was a unitary construct with invariable meaning or a complex of relatively independent dimensions with different implications and different empirical correlates. Twenty-nine indices of achievement motivation were subjected to factor analysis. Six factors were identified: Academic Motivation and Efficiency (the only factor highly predictive of actual academic performance), Wish-Fulfillment Motivation, Nonacademic Achievement Motivation, Self-Satisfaction, External Pressure to Achieve, and Imputed Generalized Motivation without Attendant Effort. In addition to demonstrating that there are several varieties of achievement motivation, the results also indicated that: tests and test items differed widely in their ability to measure any of the several facets of this complex variable, identical responses could have very different meanings for different subjects, self-rating procedures excelled projective methods in predicting actual academic performance, and the projectives were either loaded on wish-fulfillment or error factors.

REFERENCES

Atkinson, J. W. *Motives in fantasy, action, and society: A method of assessment and study.* Princeton: Van Nostrand, 1958.

Atkinson, J. W. Personality dynamics. *Annu. Rev. Psychol.*, 1960, **11**, 255–290.

Bendig, A. W. Manifest anxiety and projective and objective measures of need achievement. *J. consult. Psychol.*, 1957, **21**, 354.

Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257–261.

Birney, R. C. The achievement motive and task performance. *J. abnorm. soc. Psychol.*, 1958, **56**, 133–135.

Blake, R. R., & Mouton, J. S. Personality *Annu. Rev. Psychol.*, 1959, **10**, 203–232.

Fruchter, B. *Introduction to factor analysis.* New York: Van Nostrand, 1954.

Himelstein, P., Eschenbach, A. E., & Carp, A. Interrelationships among three measures of need achievement. *J. consult. Psychol.*, 1958, **22**, 451–452.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, **23**, 187–200.

Lazarus, R. S., Backer, R. W., Broverman, D., & Mayer, J. Personality and psychological stress. *J. Pers.*, 1957, **25**, 559–577.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Melikian, L. H. The relationship between Edwards' and McClelland's measures of achievement motivation. *J. consult. Psychol.*, 1958, **22**, 296–298.

Taylor, A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

Thurstone, L. L. *Multiple factor analysis.* Chicago: Univer. Chicago Press, 1947.

Voegl, W., Baker, R. W., & Lazarus, R. S. The roles of motivation in psychological states. *J. abnorm. soc. Psychol.*, 1958, **56**, 105–112.

(Received December 30, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 4, 188-194

CATEGORIZATION AND RELATED VERBALIZATIONS IN DEAF AND HEARING ADOLESCENTS[1]

SOLIS L. KATES, WILLIAM W. KATES, JAMES MICHAEL, AND TERRENCE M. WALSH

Clarke School for the Deaf, Northampton, Massachusetts, and University of Massachusetts

Studies dealing with the perceptual and conceptual capacities of deaf children have suggested that children deaf from birth fail to develop the ability to deal with abstract material in a fashion comparable to hearing children (Larr, 1955; Myklebust, 1953; Oleron, 1950; Templin, 1950). Other studies, however, have shown that the differences between the thinking processes of deaf and hearing children do not correspond to the differences between the concreteness of the aphasics and the abstract functioning of normal subjects (Oleron, 1953) but correspond much more closely to the differences between a younger and older group of hearing children (Clarke School, 1940). When the level of verbalization required is within the range of vocabulary and experience of the deaf subject, deaf subjects perform as well as hearing subjects on abstract as well as less abstract tasks (Rosenstein, 1959).

There are three major purposes of this experiment. First, we are interested in investigating a specific aspect of the cognitive processes of deaf and hearing children, that of categorization. Categorization, "to render discriminably different things equivalent," forms one of the bases for all other cognitive activities (Bruner, Goodnow, & Austin, 1956, p. 1). Our second major concern is in distinguishing between the process of categorization on the one hand, and the process of verbalization on the other, and to judge the adequacy of categorization independently of the adequacy of verbalization. The process of learning may involve both learning a new verbal attribute and then learning the nonverbal category that goes with it as well as learning a verbal attribute to go with a nonverbal category already formed (Brown, 1958). Our third concern is in attempting to separate the effects of deafness from the cognitive effects of age and achievement level.

The general hypothesis of this study was: There will be no differences in categorization or verbalization found for one comparison of deaf subjects with hearing subjects of matched age and IQ that will also be found for the comparison of these deaf subjects with hearing subjects matched on school achievement and IQ.

METHOD

Subjects. One group consisted of five girls and three boys at the Clarke School for the Deaf in Northampton, Massachusetts, representing the total number of graduating seniors with very profound hearing loss. Their training was by the oral method of instruction. Their average IQ on the WAIS (both verbal and performance subscales) was 102.63; *SD*, 6.93. On the Stanford Achievement Test, their average achievement was equivalent to the 7.79 grade level. Their mean age was 18.09 years; *SD*, 0.33.

Our second group was composed of eight junior high school students, five girls and three boys. Their average Stanford Achievement level was equivalent to the 7.68 grade level. Their average IQ, based on WISC and Terman-McNemar, was 104.63; *SD*, 5.05. Their average age was 14.2 years.

Our third group included eight high school seniors, five girls and three boys.

[1] This study supported in part by Grant No. 549 from the Office of Vocational Rehabilitation, United States Department of Health, Education, and Welfare.

Their average age was 17.97 years; *SD*, 0.50. Their average IQ, based on WAIS and Terman-McNemar, was 106.38; *SD*, 5.50. These subjects were achieving at an average level for their high school class, and were expected to graduate without difficulty. The high school and junior high school students were screened carefully in consultation with their teachers and counselors, and no subject was included with any atypical personality and physical characteristics.

There were no significant differences between the deaf and the junior high school subjects on achievement level and IQ, and, similarly there were no significant differences between the deaf and the high school students on age and IQ. The size of the sample makes it necessary for this study to be regarded as exploratory and, of course, subject to replication.

Categorization and Verbalization Tests. The Goldstein-Scheerer (1941) Object Sorting Test was individually administered. This test consists of 33 everyday objects including such things as a real cigar, a real pair of pliers, a toy hammer, etc. (Rapaport, Gill, & Schafer, 1945).

Procedure. The Object Sorting Test was administered in three parts: the free, active, and compliant sections. In the free sorting section, the subject was asked to put all the objects into any number of groups and then tell why the objects in each group belonged together. In the active section, the subject was asked to put all the objects together that belonged with a preselected sample object and give his reasons for the categorization. There were eight "target" objects, the first selected by the subject and rest presented by the experimenter. The compliant section involved the presentation of 12 different preselected groupings of objects and asking each subject to give the reasons for the groupings. All three sections of the test were administered in one session and all subjects received the free, active, and compliant sections in order. Written instructions were used for the deaf subjects and the same instructions were given orally to the hearing subjects. There was no doubt on the part of the experimenters that all verbalizations of all subjects were well understood.

Scoring. A scoring system for categorizations was devised, based partially on that used by Rapaport et al. (1945, pp. 401–406). Classification of verbalizations were based upon those used by Rapaport as well as upon the criteria for types of rules of verbalization set up by Bruner et al. (1956, pp. 4–7). Two independent scorers agreed on 95% of a sample of 350 items scored independently. Copies of the complete scoring system are available upon request.

Statistical Procedure. Our statistical analysis compared separately the deaf subjects with each one of the two control groups; the question of the overall differences among the three groups was not a concern of this study.

To make some data comparable from subject to subject, proportions were calculated for measures such as adequacy of categorization and verbalization. Estimate tests of homogeneity of variance were performed and heterogeneous data were converted to *T* scores (Edwards, 1954). All variances were then found to be homogeneous.

Results and Discussion

Prediction 1. The deaf subjects will have a similar number of divisions and categories as the hearing subjects matched on achievement and IQ but significantly more categories and divisions than the hearing subjects matched on age and IQ. Furthermore, the deaf subjects will not differ in the number of failures from either hearing group. Divisions are defined as the number of groupings and/or single objects that the 33 objects are divided into in the free sorting section. The number of categories is equal to the number of groupings with two or more nonidentical objects included. A failure consists of an object that is not grouped with another object or of an absolute identity class such as two sugar cubes.

Results. The results, in general, support this prediction. Deaf subjects have the same number of divisions, categories, and failures as both groups of hearing subjects (Table 1), and thus appear as effective as hearing subjects in the two control groups in reducing the complexity of this experimental world.

Prediction 2. The deaf subjects will categorize as adequately but will not

TABLE 1

t Test Results for Comparisons between Deaf and Control Groups Matched on Age and IQ (C_1) and on Stanford Achievement and IQ (C_2)

Comparison	Group	*M*	*SD*	*t*[a]
Divisions	C_1	8.75	1.79	0.298
	Deaf	9.00	1.32	
	C_2	7.88	1.60	1.460
Categories	C_1	7.13	1.05	0.182
	Deaf	7.25	1.39	
	C_2	6.50	1.22	1.070
Failures	C_1	2.13	1.97	0.00
	Deaf	2.13	0.93	
	C_2	2.25	1.92	0.155
Concrete verbalization[b]	C_1	0.080	0.078	1.418
	Deaf	0.158	0.123	
	C_2	0.212	0.115	0.857
Functional verbalization[b]	C_1	0.214	0.222	0.404
	Deaf	0.176	0.113	
	C_2	0.275	0.123	1.571
Category width with failures	C_1	5.00	1.16	2.239*
	Deaf	3.65	1.10	
	C_2	4.13	0.72	0.968
Category width without failures	C_1	5.20	1.03	2.482*
	Deaf	3.79	1.09	
	C_2	4.51	0.93	1.324
Formal verbalizations used inadequately[c]	C_1	52.33	11.07	0.031
	Deaf	52.18	6.15	
	C_2	47.87	5.88	1.340
Inclusive verbalization[d]	C_1	47.35	9.18	0.898
	Deaf	52.33	11.45	
	C_2	50.37	7.42	0.380

[a] The *t* value is the result of the comparison made of the deaf group mean with the control group mean just opposite the corresponding *t* value.

[b] Raw scores converted to proportion of total number of verbalizations.

[c] Proportion of formal inadequate to total inadequate converted to *T* scores.

[d] Proportion of inclusive verbalization to total number of verbalizations converted to *T* scores.

* Significant at $p < .05$ level.

have as many adequate verbalizations as the hearing subjects matched on age and IQ. Furthermore, the deaf subjects will have the same number of adequate verbalizations as the group matched on achievement and IQ. An adequate categorization is one in which all the objects included are relevant to each other independent of the verbalization, and there are no relevant objects present in the sample that are not included in the grouping. An adequate verbalization is defined as one that covers correctly the group sorted. An inadequate verbalization may be false, or may not take into account all the objects included in the categorization, or may take into account all objects included in the categorization but pertain to objects outside the categorization for which the verbalization is given.

Results. The results support this prediction. Analyses of variance performed on the proportion of adequate categorizations to the opportunities for categorization show that there are no differences between the deaf subjects and either of the two hearing groups (Tables 2 and 3). When the task of grouping equivalent objects together is judged independently of the verbalization of the reason for that grouping, deaf subjects perform as well as hearing subjects of their own age and IQ. Furthermore, the deaf subjects have significantly fewer adequate verbalizations than the hearing group matched on age and IQ ($F = 6.58$, $p < .05$) but there are no differences between the deaf group and the hearing group matched on achievement and IQ (Tables 2 and 3). Thus, in terms of adequacy of verbalization, the deaf subjects seem to resemble the younger hearing subjects.

Prediction 3. The deaf subjects will have a greater number of adequate categorizations accompanied by inadequate verbalizations than the control group matched on age and IQ while not differing from the hearing group matched on achievement and IQ.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE DEAF GROUP AND THE TWO CONTROL GROUPS

Variable	Group	*M*	*SD*
Adequacy of categorization	Deaf	0.713[a]	0.200
	C_1[b]	0.755	0.187
	C_2[c]	0.674	0.149
Adequacy of verbalization	Deaf	0.624[d]	0.224
	C_1	0.774	0.185
	C_2	0.582	0.197
Adequate categorization—inadequate verbalization	Deaf	52.48[e]	8.86
	C_1	45.32	6.54
	C_2	52.83	8.25
Formal generic verbalization	Deaf	52.70[e]	8.92
	C_1	51.99	9.54
	C_2	45.37	8.95
Total formal verbalization	Deaf	51.55[e]	8.90
	C_1	52.53	10.40
	C_2	45.82	8.08
Proneness to change	Deaf	48.63[e]	8.56
	C_1	52.05	9.80
	C_2	50.65	7.74

[a] Proportion of the number of adequate categories to total opportunities for categorisation.

[b] C_1 is the control group matched with the deaf on age and IQ.

[c] C_2 is the control group matched with the deaf on Stanford Achievement and IQ.

[d] Proportion of the number of adequate verbalizations to the number of categorizations.

[e] Converted to *T* scores.

Results. The analyses were conducted on standard scores converted to *T* scores (Tables 2 and 3). The results indicate that the deaf subjects have a significantly greater number of adequate categorizations accompanied by inadequate verbalizations than the control subjects matched on age and IQ ($F = 5.47$; $p < .05$). But there were no differences between the deaf subjects and those matched with the deaf on achievement and IQ. The results support our prediction. It may be pointed out that the younger control group had significantly more ade-

TABLE 3

ANALYSIS OF VARIANCE RESULTS FOR COMPARISONS BETWEEN DEAF AND THE CONTROL GROUP MATCHED ON AGE AND IQ (C_1) AND ON STANFORD ACHIEVEMENT AND IQ (C_2)

Variable	*df*	*SS*	*MS*	*F*
Adequacy of categorization[a]				
C_1 Between treatment groups	1	0.014	0.014	0.280
Error	14	0.699	0.050	
C_2 Between treatment groups	1	0.013	0.013	0.317
Error	14	0.573	0.041	
Adequacy of verbalization[b]				
C_1 Between	1	0.270	0.270	6.585*
Error	14	0.573	0.041	
C_2 Between	1	0.021	0.021	0.438
Error	14	0.677	0.048	
Adequate categorization—inadequate verbalization[c]				
C_1 Between	1	410.06	410.06	5.474*
Error	14	1048.76	74.91	
C_2 Between	1	0.96	0.96	0.009
Error	14	1544.34	110.31	
Formal generic verbalization[d]				
C_1 Between	1	5.98	5.98	0.003
Error	14	3064.02	218.86	
C_2 Between	1	645.55	645.55	3.417
Error	14	2644.15	188.82	
Total formal verbalization[e]				
C_1 Between	1	11.59	11.59	0.619
Error	14	2622.46	187.32	
C_2 Between	1	355.90	355.90	3.078
Error	14	1755.12	125.37	
Proneness to Change[c]				
C_1 Between	1	93.61	93.61	0.785
Error	14	1670.42	119.32	
C_2 Between	1	32.46	32.46	0.382
Error	14	1190.53	85.05	

[a] Raw scores converted to proportion of adequate categorizations to total number of categorisations.

[b] Raw scores converted to proportion of adequate verbalizations to total number of verbalisations (number of divisions plus 8 active plus 12 compliant).

[c] Raw scores converted to *T* scores.

[d] Proportions of formal generic verbalisations to total number of verbalisations converted to *T* scores.

[e] Proportions of total formal verbalisations to total number of verbalizations converted to *T* scores.

* Significant at $p < .05$ level.

quate sortings accompanied by inadequate verbalizations than the older control group. Our results suggest that deafness holds back the process of correctly relating verbal attributes to nonverbal categories but that this retardation is similar to that shown by younger hearing subjects when they are compared to older hearing subjects.

Prediction 4. The deaf subjects will have more functional, more total concrete, fewer formal, fewer formal generic verbalizations than the control group matched on age and IQ but the same number as the control group matched on achievement and IQ. A functional verbalization specifies a rule of categorization based upon the sharing of a common function. Total concrete verbalizations include the syncretistic, fabulatory, chain, symbolic, split narrow verbalization as defined by Rapaport et al. (1945), and formal specific verbalization type. Formal verbalizations are defined as verbalizations specifying a rule of grouping based on the sharing of an intrinsic property ("all red," or "all kitchen utensils") among the objects in the group (Bruner et al., 1956). A formal generic verbalization specifies a rule of grouping based upon the sharing of highly complex and class-like attributes ("all metal").

Results. There are no differences between the deaf and either of the hearing control groups with respect to any of the verbalization types discussed above (Tables 1, 2, and 3). This prediction is not supported. However, the group matched with the deaf on achievement and IQ had significantly more "total concrete" verbalizations than the older control group. This result indicates that the test does differentiate, at least in regard to concrete verbalizations, levels of verbalization due to age and achievement differences in hearing subjects. On the whole these concrete verbalizations represent an inability to utilize a conjunctive rule of grouping to explain a categorization. Further, they can also be described as "private" verbalizations (McGaughran & Moran, 1956). Deaf subjects on these tasks are no more concrete verbally, do not demonstrate any greater inability to utilize conjunctive rules of grouping, and utilize no more private verbalizations than the two control groups.

There is little evidence and theory concerning the following issues, and they are stated in the form of a question.

Are there any differences between the deaf subjects and either control group in proneness to change, in category width, in number of formal verbalizations used inadequately, and in the number of too inclusive verbalizations? Proneness to change is measured by the exclusion of objects in the active sorting which were included with these target objects in free sorting or by the inclusion, with the target objects in the active sorting, of objects not included with them in the free sorting. Category width is measured by the number of objects included with the target objects in the active section, both counting and not counting failures.

The data were converted to *T* scores and analysis showed that there are no differences between deaf subjects and either of the two control groups in overall proneness to spontaneous change (Tables 2 and 3). These results are in contrast to other results which have suggested that deaf subjects are more rigid than hearing children (McAndrew, 1948). The question of the rigidity of the deaf is not a settled one, since our evidence points to no difference between groups when spontaneous behavior is examined.

Our results show that the deaf sub-

jects have narrower categories than the control group matched on age and IQ but do not differ in category width from the hearing subjects matched on achievement and IQ (Table 1). The use of the narrower categories by the deaf subjects who, at the same time have the same number of categories as the hearing groups, may indicate a greater difficulty in organizing the world and in generalizing rules of grouping to a wide array of objects. This, of course, would also be true of the younger group of hearing subjects from whom the deaf did not differ in category width. However, with respect to this small array of familiar objects, the use of narrower categories might be more appropriate and more predictive than the use of wider categories. The problem of the relation of deafness to category width seems to be one that deserves future attention.

The two proportions of formal inadequate verbalizations and of inclusive verbalizations to the total number of verbalizations, respectively, were converted to *T* scores and were compared for the three groups. The deaf did not differ significantly from either control group in formal inadequate verbalizations and in inclusive verbalizations, respectively (Table 1).

Summary

This study has investigated the categorization and verbalization processes of 24 adolescents, 8 of whom were profoundly deaf. The remaining subjects comprised two hearing groups, one group matched with the deaf subjects on age and IQ and the other group matched on Stanford Achievement and IQ. The Goldstein-Scheerer Object Sorting Test was administered and the deaf subjects were separately compared with each of the two control groups.

Our conclusions are as follows:

1. Deaf subjects categorize, on this task, as adequately as the hearing subjects. They have the same number of divisions, failures, and categories as the hearing subjects. On categorization tasks judged independently of verbalization, deaf subjects perform as well as hearing adolescents.

2. Deaf subjects have more inadequate verbalizations than hearing subjects and more adequate categorizations accompanied by inadequate verbalizations than hearing subjects matched on age and IQ. These differences are not true for the comparison of the deaf subjects and the hearing subjects matched on achievement and IQ. It is proposed that deafness retards learning of the correct relation between the proper verbal attribute and the nonverbal category but that this retardation lies along a normal developmental progression.

3. Deaf subjects do not differ from hearing subjects in the type of verbalization or in the developmental level of the verbalizations used, or in their spontaneous changes of categorization throughout the test.

4. Deaf subjects have narrower categories than hearing adolescents of their own age and IQ though they do not differ from hearing subjects of the same achievement and IQ.

5. There were no differences found between deaf and hearing subjects that indicated a definite shift in categorization ability distinct from the effects of age and achievement.

REFERENCES

Brown, R. *Words and things*. Glencoe, Ill.: Free Press, 1958.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

Clarke School for the Deaf, Clarence W. Barron Research Department, Psychological Division. Studies in the psychol-

ogy of the deaf. *Psychol. Monogr.*, 1940, **52**(1, Whole No. 232).

EDWARDS, A. L., *Statistical methods for the behavioral sciences.* New York: Rinehart, 1954.

GOLDSTEIN, K., & SCHEERER, M., Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

LARR, A. L. An experimental investigation of the perceptual and conceptual abilities of children in residential schools for the deaf. Unpublished doctoral dissertation, Syracuse University, 1955.

MCANDREW, H. Rigidity and isolation: A study of the deaf and the blind. *J. abnorm. soc. Psychol.*, 1948, **43**, 476–494.

MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *J. abnorm. soc. Psychol.*, 1956, **52**, 43–51.

MYKLEBUST, R., Towards a new understanding of the deaf child. *Amer. Ann. Deaf*, 1953, **98**, 345–357.

OLERON, P. A study of the intelligence of the deaf. *Amer. Ann. Deaf*, 1950, **95**, 179–195.

OLERON, P. Conceptual thinking of the deaf. *Amer. Ann. Deaf*, 1953, **46**, 299–315.

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing.* Vol. 1. Chicago: Year Book Publishers, 1945.

ROSENSTEIN, J. Perception and cognition in deaf children. Unpublished doctoral dissertation, Washington University, 1959.

TEMPLIN, M. The development of reasoning in children with normal and defective hearing. Minneapolis: *Univer. Minnesota Press*, 1950.

(Received September 6, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 4, 195–198

CROSS-VALIDATION GROUPS, EXTREME GROUPS, AND THE PREDICTION OF ACADEMIC ACHIEVEMENT

LEONARD J. KAZMIER[1]
University of Notre Dame

In two parallel studies, scores on the 15 personality variables of the Edwards Personal Preference Schedule (EPPS) (Edwards, 1954) have been compared for over- and underachieving college students (Gebhart & Hoyt, 1958; Krug, 1959). In both studies, extreme over- and underachievers were selected from a larger sample of male college students, so that the middle range of achievers was eliminated from the study. Accordingly, analyses of variance were performed to test for the significance of differences between the extreme groups, but validity coefficients were not computed. Though the scales that were found to discriminate between subjects who over- or underachieved in relation to measured aptitude were not identical, the results of the two studies did agree to a considerable extent. In the Gebhart and Hoyt study, overachievers scored higher than underachievers on the Achievement, Order, Intraception, and Consistency scales, and significantly lower on the Nurturance, Affiliation, and Change scales. In the study by Krug, overachievers scored significantly higher on the Achievement, Order, and Endurance scales and significantly lower on Affiliation and Heterosexuality. The .05 level was defined as significant in both studies.

There has come to be considerable concern about the widespread use of extreme groups in hypothesis testing. As McNemar has indicated (1960), such tests of significance are misleading and may represent only a small degree of relationship between the two variables. The primary purpose of the present study was to investigate the relationship between the EPPS scales and a criterion of academic achievement when scores of all subjects are utilized, not just those of selected extreme groups. The correlation technique was used throughout because it provides a measure of the degree of relationship as well as its significance. To ascertain that the measured relationships of the several scales with the criterion are not due to chance factors, a cross-validation group was included in the study.

Procedure

The sample consisted of 140 students enrolled in two sections of introductory psychology at Ohio State University during the Spring Quarter of 1960. Both sections were taught by the same instructor, the author of the present report. For use as a predictor, six paragraphs of material, including a total of 36 questions on the contents thereof, were chosen from old forms of the Ohio State Psychological Examination (Buros, 1959, pp. 494–495). The criterion for success in the psychology course was the total number of points accumulated at the end of the quarter from three departmental examinations and four class quizzes.

In order to achieve comparability of the experimental and cross-validation groups, all class cards for the subjects were segregated by sex, and then categorized according to college year. Within the college year the cards were arranged in order of percentile scores on the Ohio State Psychological Examination. Seven students for whom these scores were not available were arbitrarily placed at the end of their respective college-year groups, after the highest percentile scores. Then the overall sample was subdivided into two groups of 70 subjects each by alternatively putting one card in the experimental group and the following card in the cross-validation group. Comparability of the two groups in terms of

[1] Research conducted while the author was at Ohio State University.

their standing on the reading comprehension test, the criterion, and the 15 EPPS scales was ascertained.

The correlation of each of the 15 scales with the criterion of success was determined for each of the groups and for the entire sample, and the Wherry-Doolittle test selection technique (Garrett, 1958, pp. 426–440) was then applied to the experimental group of subjects in order to determine the best selection and weighting of EPPS scales for maximum predictiveness. The applicability of this result to the cross-validation group was investigated by computing a composite r for that group by using the scales and weights derived from the experimental group.

Results

The split-half reliability coefficient (odd vs. even) for the reading comprehension test is .813, while the coefficient of correlation between this predictor and the criterion of total points in the psychology course is .593, which is significant at the .001 level for the 140 subjects involved in this study.

TABLE 1

Comparisons between the Experimental and Cross-Validation Groups

Measure	Experimental Group (N = 70)		Cross-Validation Group (N = 70)		
	M	SD	M	SD	F
Points in psychology course	259.00	39.25	260.10	36.05	—[a]
Predictor Score	17.18	6.22	16.84	5.98	—
EPPS Scales:					
Achievement	13.67	3.94	14.23	4.75	—
Deference	11.41	3.71	12.21	3.67	1.04
Order	9.44	4.17	10.67	3.90	3.24
Exhibition	14.34	3.33	14.34	3.65	—
Autonomy	13.86	4.20	13.66	5.04	—
Affiliation	16.20	4.38	16.57	4.49	—
Intraception	17.26	4.88	17.34	4.78	—
Succorance	11.38	5.25	11.00	4.13	—
Dominance	15.58	4.68	14.41	5.23	1.95
Abasement	14.31	4.99	14.70	5.67	—
Nurturance	15.66	5.16	16.18	4.81	—
Change	17.57	4.78	16.90	4.55	—
Endurance	13.63	5.67	13.08	5.27	—
Heterosexuality	14.97	5.08	15.08	5.58	—
Aggressiveness	11.00	4.62	10.44	4.54	—

[a] $F < 1.00$.

TABLE 2

Correlation of the EPPS Scales with the Criterion of Success in Psychology

Scale	Experimental Group (N = 70)	Cross Validation Group (N = 70)	Total Sample (N = 140)
Achievement	.113	.191	.153*
Deference	.013	−.109	−.043
Order	.029	−.214*	−.080
Exhibition	.027	.157	.092
Autonomy	.073	−.040	.013
Affiliation	−.139	−.161	−.149*
Intraception	.022	.099	.059
Succorance	−.205*	.078	−.087
Dominance	.209*	.226*	.213**
Abasement	−.054	−.068	−.060
Nurturance	−.167	−.163	−.164*
Change	−.244**	.035	−.114
Endurance	.088	−.061	.018
Heterosexuality	.192	.037	.114
Aggressiveness	.115	−.034	.043

* $p < .10$.
** $p < .05$.

That the procedure used to divide the subjects into experimental and cross-validation groups did result in comparability of the groups in many important respects is indicated by the data in Table 1, which presents the mean scores and standard deviations for the two groups on the reading comprehension test, the criterion, and the 15 EPPS scales. Not one of the obtained F ratios is significant at the .05 level.

The correlation coefficients between the criterion of success in the psychology course and each of the 15 scales of the EPPS are indicated in Table 2 for the experimental and cross-validation groups as well as for the total sample. For the experimental group, the correlation coefficients for Succorance and Dominance were significant at the .10 level, while that for Change was significant at the .05 level. The cross-validation group yielded a somewhat different pattern of correlations, re-

sulting in significance at the .10 level for the Order and Dominance scales. For the total sample of subjects, the correlation coefficients for Achievement, Affiliation, and Nurturance were significant at the .10 level while that for Dominance was significant at the .05 level.

It would appear that the desired stability of relationship of the EPPS scales with the criterion is lacking in this study. For example, of the three scales whose relationship with the criterion was significant at the .10 level or better in the experimental group, only one scale "held up" for the cross-validation group, while the algebraic signs of the correlation coefficients for the other two groups actually were reversed. In order to verify this conclusion statistically, the Wherry-Doolittle technique was utilized to ascertain the best selection of scales for the experimental group and the appropriate beta weights to be used in conjunction with each scale. Stability of the scales is then indicated by the extent to which these beta weights apply for use with the cross-validation group. As indicated by Table 3, for the experimental group it was found that the use of the Change, Succorance, and Heterosexuality scales results in the maximal coefficient of multiple correlation corrected for shrinkage. Addition of the fourth best scale, Abasement, results in a decrease in the magnitude of the correlation coefficient.

The beta weights obtained by the Wherry-Doolittle test selection method are −.252 for Change, −.233 for Succorance, and .181 for Heterosexuality. Utilizing these weights derived for the experimental group and the correlation coefficients of the chosen scales with the criterion in the cross-validation group results in a composite coefficient of correlation for the cross-validation group of −.058.

TABLE 3

BEST SELECTION OF EPPS SCALES FOR THE EXPERIMENTAL GROUP BY THE WHERRY-DOOLITTLE METHOD

Scales	R
Change	.244
Change + Succorance	.313
Change + Succorance + Heterosexuality	.345
Change + Succorance + Heterosexuality + Abasement	.331

Thus, the validity evidenced for the EPPS scales in the experimental group did not at all hold up in the cross-validation group, in that the scales and weights that resulted in a multiple R of .345 for the experimental group yielded a slightly negative composite coefficient for the second group.

SUMMARY

Previous studies have reported predictiveness for certain scales of the Edwards Personal Preference Schedule (EPPS) in regard to academic over- and underachievement when the analysis of variance technique was applied to selected extreme groups. The purpose of the present study was to test out the EPPS in a situation in which no subjects are eliminated from the study, the correlation technique is used, and both experimental and cross-validation groups are included.

The selection of the best EPPS scales through the use of the Wherry-Doolittle test selection technique resulted in a coefficient of multiple correlation of .345 for the experimental group. However, when these scales were used for the cross-validation group, the value of the correlation coefficient degenerated to −.058.

The results of the study do not necessarily indicate a lack of usefulness for the EPPS, but they do highlight

the shortcomings of many of the research studies which have reported success in prediction with the EPPS scales.

REFERENCES

BUROS, O. K. *The fifth mental measurements yearbook*. Highland Park, New Jersey: Gryphon, 1959.

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1954.

GARRETT, H. E. *Statistics in psychology and education*. New York: Longmans, Green, 1958.

GEBHART, G. G., & HOYT, D. P. Personality needs of under- and overachieving freshmen. *J. appl. Psychol.*, 1958, **42**, 125–128.

KRUG, R. E. Over- and underachievement and the Edwards Personal Preference Schedule. *J. appl. Psychol.*, 1959, **43**, 133–136.

MCNEMAR, Q. At random: Sense and nonsense. *Amer. Psychologist*, 1960, **15**, 295–300.

(Received November 29, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 4, 199-200

A NOTE ON RELATIONSHIPS BETWEEN STUDENT AND PARENT MINNESOTA TEACHER ATTITUDE INVENTORY SCORES

ROBERT E. CAMPBELL AND JOHN E. HORROCKS
Ohio State University

Influences contributing to the formation of attitudes toward teaching and pupil-teacher relationships are complex. They are also difficult to identify with any degree of comprehensiveness or accuracy. Riccio (1959), in a survey and study of the literature of teacher attitudes identified a number of related variables. The variables included personality traits, experiences, values, and, to some extent, academic aptitude. To the best of the writers' knowledge, an overlooked variable has been the role of parents in helping form the professional attitudes of those of their children who become teachers. The present study deals with possible relationships between professional attitudes of teachers-in-training and those of their parents. Three questions were considered: How similar are the attitudes of students and their parents? Does the student tend to identify attitudinally more with one parent than with the other? What is the relationship between the attitudes of a student's parents?

Procedure

The Minnesota Teacher Attitude Inventory (MTAI) (Cook, Leeds, & Callis, 1952) was administered to 127 Ohio State University College of Education students and their parents. The students were mainly sophomores enrolled in an educational psychology course. Both students and parents volunteered to participate in the study as a partial course requirement. The MTAI was administered to the students in a classroom session at the beginning of the course. A month later inventories and an explanatory letter were sent to parents. The letter stated that the investigators were interested in parental attitudes toward education. The parents were requested to answer their questionnaires separately and to refrain from showing their answers to each other. Returned answer sheets were carefully screened for misinterpretation of directions.

Results and Discussion

Table 1 shows the relationship between the MTAI scores of students and those of their parents. The data suggest three trends: the moderate positive (.45) agreement between parents' attitudes; the similarity of attitudes between students and their parents; and though tenuously, the students' attitudinal identification with the parent of the same sex. Daughters and mothers present a higher relationship (.37) than did daughters and fathers (.15).[1] Similarly the son-father correlation (.33) was higher than the son-mother correlation (.19). Although these differences were not statistically reliable, trends in these directions were not unexpected in view of similar results reported in studies made in other areas of parent-child identification. Father-son and mother-daughter identifications were noted by Gray and Klaus (1956) in their study of values, and by Henderson (1958) in his study of interests. Lazowick (1955), using an S-R approach, has also reported on the nature of identification.

Table 2 presents the means and standard deviations for the categories of groups used. Although the variance

[1] The difference between the correlations was not statistically significant at the .05 level of confidence.

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN STUDENTS' AND PARENTS' MTAI SCORES

Group	*r*	*N*
Students-fathers	.285*	127
Students-mothers	.334*	127
Mothers-fathers	.449*	127
Sons-fathers	.330*	39
Sons-mothers	.189**	39
Daughters-fathers	.149**	88
Daughters-mothers	.369*	88

* Significant at .01 level of confidence.
** Significant at .05 level.

TABLE 2

MEANS AND STANDARD DEVIATIONS OF MTAI SCORES FOR PARENTS AND STUDENTS

Group	*N*	*M*	*SD*
All students	127	30.4	28.4
Mothers	127	1.0	34.5
Fathers	127	−14.3	29.0
Daughters	88	35.4	28.2
Sons	39	18.9	25.2

is homogeneous for all groups, the means are significantly different. Since the MTAI may be interpreted on an authoritarianism-permissiveness continuum (high positive scores represent permissiveness in teacher-pupil relationships whereas low positive or negative scores present authoritarian attitudes) it is possible to rank daughters as most permissive with sons, mothers, and fathers following in that order. It would appear that an age as well as a sex variable may have been operating here.

SUMMARY AND CONCLUSIONS

Similarities of teaching attitudes of 127 teachers-in-training and their parents were investigated by means of the Minnesota Teacher Attitude Inventory.

The results suggest the following conclusions: (*a*) parents are in relatively moderate agreement, (*b*) a fair positive relationship exists between the attitudes of students and those of their parents, (*c*) students tend to identify attitudinally with the parent of the same sex, (*d*) daughters are the most permissive in their concept of teacher-pupil relationships while sons are less so, and (*e*) mothers and fathers tend to be more authoritarian in their MTAI responses than do their children of either sex.

REFERENCES

COOK, W. W., LEEDS, C. H., & CALLIS, R. Minnesota Teacher Attitude Inventory manual. New York: Psychological Corporation, 1952.

GRAY, S. W., & KLAUS, R. Assessment of parental identification. *Genet. psychol. Monogr.*, 1956, **54**, 87–109.

HENDERSON, H. L. The relationship between interests of fathers and sons and sons' identification with fathers: The relationship of the adolescent son's identification with his father to father-son interest similarity as measured by the Strong Vocational Interest Blank. Unpublished doctoral dissertation, Columbia University, 1958.

LAZOWICK, L. M. On the nature of identification, *J. abnorm. soc. Psychol.*, 1955, **51**, 175–203.

RICCIO, A. C. The relationship of selected variables to attitudes toward teaching. Unpublished doctoral dissertation, Ohio State University, 1959.

(Received December 29, 1960)

Journal of Educational Psychology
1961, Vol. 52, No. 4, 201-206

TEST ANXIETY AND THE INTELLECTUAL PERFORMANCE OF COLLEGE STUDENTS[1]

IRWIN G. SARASON[2]
University of Washington

A currently strong trend in the selection of college students is the reliance on objective tests of intelligence and aptitude. Very likely the problems associated with an increasing college age population and the heightened demands for admission to colleges are going to accelerate this trend. While the value of entrance examinations and objective measures of intellectual performance is obvious, a legitimate question would seem to be whether or not factors other than ability and aptitude affect the test performance of students. If it were the case, for example, that certain personality characteristics correlated negatively with intellectual performance measures, it would seem important to take this correlation into account in some way in evaluating applicants.

Although the problem of the possible confounding of personality with intellectual variables has been stated specifically in terms of admission to college, it is clear that this is but one aspect of the more general problem of the relationship between intelligence and personality. This is so because all students in most school systems undergo systematic intellectual assessment, and educational decisions frequently are made on the basis of this assessment.

[1] This research was supported by a grant from the Agnes H. Anderson Research Fund.

[2] The writer is indebted to Paul Horst and the staff of the Counseling and Testing Service of the University of Washington for their cooperation in making available these data.

The present research investigated the relationship between intellectual and personality measures for male and female freshman and sophomore college students. Thirteen intellectual measures were available for each student, of which six were high school grade point averages in a variety of fields, and seven were objective tests included in an entrance battery. Six personality variables were correlated with each of these 13 intellectual measures. Three of these six personality variables were measures of what might be called general anxiety, i.e., anxiety as expressed or experienced in a variety of situations. One other measure tapped the student's feelings of hostility in a variety of situations and another measure was designed to assess the subject's level of defensiveness. The sixth personality variable studied, and the one which it was felt would most strongly relate to intellectual performance, was test anxiety, i.e., anxiety experienced specifically in testing situations. The prediction of significant relationships between test anxiety and intellectual performance was based on the following formulation.

Anxiety responses are aroused by personally threatening conditions in the environment. These responses interfere with ongoing task-relevant activity and lead to a lowering of performance level. Although the exact nature of the interfering anxiety responses is not known, it seems likely that they include both heightened

physiological activity (e.g., heart rate, palmar sweating) and self-deprecating ruminations (e.g., "I can't pass this test," "I wonder how the other students are doing"). While these anxiety responses are assumed to be aroused under stressful conditions, their evocation would not necessarily be expected under neutral conditions (Sarason, 1960).

Within this view of anxiety, anxiety scales or questionnaires may be considered to measure the likelihood of emission of anxiety responses under conditions of personal threat. Furthermore the closer the content of the anxiety scale is to the situation in which the subject's performance is measured the better predictor of reaction to threat will the anxiety scale be. Since the test anxiety scale used in this study contains items dealing with the subject's reactions to testing situations, it was expected that it would most meaningfully relate to performance. Because of the importance of college aptitude tests to students it was predicted that test anxiety would correlate negatively with the seven entrance examination tests. The correlations of test anxiety with the six high school grade point averages were expected to be of a lower order than those with the entrance test battery. This latter prediction was made because of the possibility that while highly anxious students might not be able to prepare for and cope adequately with entrance and aptitude tests, they well might be able to reduce their anxiety levels over a semester or school year in course work.

Method

Subjects. The subjects were 326 male and 412 female students enrolled in introductory psychology and introductory sociology courses at the University of Washington. The Autobiographical Survey containing the personality measures (Sarason, 1958b) was administered during regular class sessions. The aptitude measures were obtained prior to the subjects' entering the University of Washington. Thus the personality and intellectual measures were not obtained contemporaneously.

Materials. The form of the Autobiographical Survey used in this research consists of six true-false scales: Test Anxiety, (TA), 17 items; General Anxiety (GA), 18 items; Lack of Protection (LP), 27 items; Hostility (H), 22 items; Need for Achievement (NA), 31 items; Defensiveness (D), 24 items. Descriptions of the scales have been presented previously (Sarason 1958b).[8]

The 13 intellectual measures employed are routinely gathered for entering University of Washington students. The measures are:

1. High school English average
2. High school mathematics average
3. High school foreign language average
4. High school social studies average
5. High school natural science average
6. High school electives average
7. Verbal score on the Guilford-Zimmerman Aptitude Survey
8. Mechanical Knowledge score on the Guilford-Zimmerman Aptitude Survey
9. Cooperative English Usage score
10. Cooperative English Spelling score
11. Cooperative Mathematics I score
12. Cooperative Social Studies II score
13. ACE (1948) Q score

Results

The results for the male and female subjects will be presented first in terms of the correlations between the intellectual and personality measures and then in terms of the intercorrelations of the Autobiographical Survey scores.

Table 1 contains the correlations between the 6 personality and 13 intellectual variables for the 326 male subjects. The table also contains the means for the 13 intellectual measures. Although GA correlated negatively and significantly with 4 intellectual measures and NA correlated negatively and significantly with 5 of these

[8] Copies of the Autobiographical Survey may be obtained from the author.

TABLE 1

CORRELATIONS OF SIX AUTOBIOGRAPHICAL SURVEY SCORES WITH THIRTEEN INTELLECTUAL MEASURES FOR 326 MALES

Measure	TA	GA	LP	H	NA	D	Mean
High school English average	−09	−04	00	−06	−05	06	2.82
High school mathematics average	−18	−03	−01	−03	−09	08	2.61
Foreign language average	−09	03	09	−03	−07	00	2.57
Social science average	−17	−04	−01	−08	−01	07	2.95
Natural science average	−15	−07	02	−07	−06	07	2.78
Elective average	−14	−07	−04	−12	−09	07	3.06
G-Z Verbal	−23	−03	00	05	−18	00	27.18
G-Z Mechanical Knowledge	−13	−11	−07	00	−16	03	21.65
Cooperative English Usage	−27	−12	−02	−06	−14	11	86.61
Cooperative English Spelling	−18	−05	−04	00	−11	03	16.17
Cooperative Mathematics I	−13	−10	−03	00	00	11	23.71
Cooperative Social Studies II	−23	−11	−01	06	−13	08	18.45
ACE Q score (1948)	−30	−17	−06	−01	−06	10	46.96

Note.—r = .123 at .05 level; r = .148 at .01 level.

measures, it is clear that TA shows the most consistent set of correlations. For 11 of the 13 intellectual measures there were significant negative correlations with TA.

A comparable set of correlations for 412 females (Table 2) also shows TA to correlate negatively and significantly with 11 of the 13 intellectual measures, although the same 11 variables were not significant for both males and females. The only other personality characteristic found to be related to intelligence was H which correlated positively with four intellectual measures.

For both the male and female groups, comparisons were made of the correlations obtained between TA and high school grade point averages and TA and entrance tests. For the purposes of this comparison the six TA grade point averages were compared with six TA entrance test scores. Not considered were the male and female TA–Mechanical Knowledge correlations. This omission was based on the great differences between male and female means on this test. The comparisons made indicated stronger negative correlations between TA and the six entrance tests than between TA and the six high school grade point averages. This result was significant for both males and females ($p < .05$ and $p < .02$, respectively).

Intercorrelations between the six Autobiographical Survey measures and the means for these measures are presented in Tables 3 and 4, for males and females, respectively. In general the patterns of intercorrelations for males and females are quite comparable. For both sexes the D measure is

TABLE 2

CORRELATIONS OF SIX AUTOBIOGRAPHICAL SURVEY SCORES WITH THIRTEEN INTELLECTUAL MEASURES FOR 412 FEMALES

Measure	TA	GA	LP	H	NA	D	Mean
High school English average	−12	02	01	02	00	−02	3.30
High school mathematics average	−12	03	06	01	−01	−07	2.81
Foreign language average	−14	10	05	07	−05	−07	3.08
Social science average	−13	02	03	05	−02	−04	3.24
Natural science average	−19	−01	05	06	−04	00	3.06
Elective average	−06	00	01	−01	05	−01	3.33
G-Z Verbal	−22	00	07	13	−07	−03	27.55
G-Z Mechanical Knowledge	−06	−09	−04	10	−05	−03	4.51
Cooperative English Usage	−25	00	06	09	−03	−02	101.75
Cooperative English Spelling	−16	03	00	11	−04	−03	20.47
Cooperative Mathematics I	−21	−04	02	09	−01	00	17.15
Cooperative Social Studies II	−20	−05	04	11	−06	−01	16.43
ACE Q score (1948)	−18	−04	.01	07	06	−01	43.88

Note.—r = .098 at .05 level; r = .128 at .01 level.

TABLE 3

INTERCORRELATIONS OF SIX AUTOBIOGRAPHICAL SURVEY SCORES FOR 326 MALES

Scores	TA	GA	LP	H	NA	D	Mean
TA		53	35	13	34	−50	4.97
GA			58	23	34	−64	5.57
LP				33	37	−56	6.32
H					20	−40	8.86
NA						−35	11.57
D							12.50

Note.—r = .113 at .05 level; r = .148 at .01 level.

TABLE 4

INTERCORRELATIONS OF SIX AUTOBIOGRAPHICAL SURVEY SCORES FOR 412 FEMALES

Scores	TA	GA	LP	H	NA	D	Mean
TA		48	35	08	35	−45	5.64
GA			54	26	31	−60	6.40
LP				33	24	−55	7.08
H					14	−44	6.96
NA						−31	10.97
D							13.42

Note.—r = .098 at .05 level; r = .128 at .01 level.

negatively related to each of the other Autobiographical Survey measures.

DISCUSSION

The present findings demonstrate that for both men and women there are significant negative relationships between anxiety reported by subjects as being experienced in testing situations (TA) and a variety of measures of intellectual performance. These results also confirm the findings of a previous study (Sarason, 1959) which showed (*a*) TA to be negatively related to many of the same intellectual variables with which the present study was concerned and (*b*) other personality measures to be unrelated to intellectual performance. In the present research none of the personality scales, except TA, related significantly to performance in any consistent manner for both male and female college students. Thus, support was given for the hypothesis that the more specific the measure of personality (in this case, anxiety) is to the situation being studied, the more consistently will the personality measure relate to performance (Sarason, 1960). It was hypothesized also that TA scores would be more negatively related to aptitude test measures than to the six high school grade point averages. The present study suggests that this well may be the case. The Need for Achievement scale might have been expected to relate negatively to intellectual performance since its items appear to reflect strong need and anxiety to achieve. There was an indication that this might be true for male students but not for female students.

Two problems then seem to be posed by the present findings of consistent negative relationships between TA and 13 intellectual measures and also of greater strength of those negative relationships involving aptitude tests as compared with high school grade point averages. First, what factors contribute to the negative relationships obtained? Second, why is test anxiety less of a detrimental factor for course grades than for aptitude tests?

With respect to the first question, it would seem prudent at present to interpret the present results as simply indications of poorer intellectual performance for high than for low test anxious students. If this is due to intellectual differences among subjects differing in TA scores, then it may be the case that high test anxiety scores simply reflect realistic concern over intellectual ability. If, however, students differing in test anxiety do not differ in intellectual ability then it would be worthwhile to determine

which psychological aspects of testing and academic situations arouse anxiety responses in high TA scorers and whether or not the conditions of either the testing situation or of anxious students can be modified in such a way as to raise anxious students' performance level. Results of experimental studies indicate that under neutral or reassuring conditions high and low anxious subjects do not differ in their performance levels but that under stress they do differ in the same direction as obtained in the present research (Sarason, 1958a; Sarason & Palola, 1960).

In view of this evidence it seems possible that the answer to the second question posed, that of the explanation of the stronger correlations with TA of aptitude test scores than of grade point averages, may be in the opportunity of the student in class to be reassured and to reassure himself. In the course of a quarter or a semester the anxious student can come to know his instructors and peers, as well as to increase his familiarity with and to overlearn the material to be learned. These possibilities are largely absent in group aptitude testing. Greater emphasis in future research on the interpersonal aspects of learning and performance situations should make clearer the implications of these findings.

Theoretically, the present study has supported the interfering response interpretation of anxiety discussed earlier. These results are consistent with a conception of the TA as an index of proneness to performance disruption under personally threatening conditions. Practically, these findings strongly suggest the need to take account of the test anxiety factor in interpreting students' intellectual performance.

One factor of importance in the present study concerns that of the population of subjects used. While TA correlated consistently with the intellectual measures used, in most cases the magnitude of the r's was not great. It seems possible that the restriction in range of intelligence inherent in a college sample may have worked against obtaining even more significant r's than were found in the present research. Future research with high school level subjects and a greater range in intellectual ability should, if this is correct, lead to larger r's than were obtained in this study.[4]

The intercorrelations of the Autobiographical Survey scales show essentially the same patterns for the women as for the men. Interestingly the present intercorrelations for college students are quite similar to the intercorrelations previously reported for hospitalized psychiatric patients (Sarason, 1958b). The only sizable differences in the correlations for the two subject populations are the r's involving the Hostility scale. For the patient population this scale correlated much more strongly with the other Autobiographical Survey measures than was true in the case of college students.

Summary

The present study correlated scores on six personality scales of the Autobiographical Survey with scores on seven aptitude test scores and six high school grade point averages. The subjects were 326 male and 412 female college students.

The results demonstrated that of the six personality scales only the Test Anxiety scale (TA) related consistently to the 13 intellectual measures. For both males and females, TA

[4] A recently completed study conducted by the writer with high school students as subjects provides support for this expectation.

correlated negatively and significantly with 11 of the 13 intellectual measures. There were indications, however, that the negative correlations with TA were stronger for aptitude test scores than for grade point averages. These results were interpreted as supporting the hypotheses that: (*a*) the closer the content of anxiety scales' items is to the situation in which subjects are to perform the more predictive will the items be, (*b*) TA scores reflect the proneness of subjects to emit interfering anxiety responses under conditions of threat in a testing situation.

Comparisons of the intercorrelation matrices of the six Autobiographical Survey scales for males and females showed quite similar patterns. The Autobiographical Survey intercorrelations for male and female college students were found to be similar to those previously reported for a neuropsychiatric population.

REFERENCES

SARASON, I. G. Effects on verbal learning of anxiety, reassurance, and meaningfulness of material. *J. exp. Psychol.*, 1958, **56**, 472–477.(a)

SARASON, I. G. Interrelationships among individual difference variables, behavior in psychotherapy, and verbal conditioning. *J. abnorm. soc. Psychol.*, 1958, **56**, 339–344.(b)

SARASON, I. G. Intellectual and personality correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1959, **59**, 272–275.

SARASON, I. G. Empirical findings and theoretical problems in the use of anxiety scales. *Psychol. Bull.*, 1960, **57**, 403–415.

SARASON, I. G., & PALOLA, E. G. The relationships of test and general anxiety, difficulty of task, and experimental instruction to performance. *J. exp. Psychol.*, 1960, **59**, 185–191.

(Received January 23, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 4, 207-214

EFFECTS OF MEMORY, EVALUATIVE, AND CREATIVE READING SETS ON TEST PERFORMANCE

E. PAUL TORRANCE AND JUDSON A. HARMON
University of Minnesota

For several years the senior author (Torrance, 1950, 1951) has been experimenting with a variety of procedures which might increase the likelihood that students in mental hygiene courses will make use of scientifically developed principles of mental health in their work. Several recent experiments (Torrance, 1960) have been motivated by the belief that the failure of educators to make more effective use of such principles results from the fact that they have not been educated to think creatively or imaginatively about course materials. It soon became evident in this experimentation that many, perhaps most, students did not have the expectation that research results or scientifically developed principles *could* be applied imaginatively in solving many of their personal and professional problems. At best they were looking for specific prescriptions.

The rather shocking realization that students enrolled in a graduate course in mental hygiene did not even expect to be able to use the knowledge they were acquiring led the senior author to try to develop ways of establishing appropriate "sets." In one study (Torrance, 1960) random halves of two classes were asked to read research articles creatively or critically. Those who read and reported on the articles creatively were later more successful in developing a new idea of their own and in responding to a test of creative applications. These results stimulated the present study, an extension of "sets" to textbook readings.

The effect of appropriate preparatory sets in facilitating learning has been well-established (Johnson, 1955, p. 77). Usually a person's activities are integrated and he does not react to every event that stimulates his sense organs. He is selective not only of what he responds to but also of what is produced and of the mental operation used in responding.

In designing the present experiment, the authors have been guided by Guilford's (1959) model of human intellect. Particular attention has been paid to his five mental operations: cognition, memory, convergent thinking, divergent thinking, and evaluation. By "cognition," Guilford means the operations involved in recognizing facts, discovery, and rediscovery. Memory involves the retention of what has been cognized. The two kinds of productive thinking, convergent and divergent, include the generation of new information from known information and remembered information. Convergent thinking leads to a conventional solution; divergent thinking leads to new and untested solutions. Evaluation refers to decisions as to the goodness, suitability, or adequacy of what we know, remember, or produce through convergent or divergent thinking. In this experiment a deliberate attempt was made to establish sets which would result in utilization of these operations.

METHOD

Subjects

The subjects were 115 students in a graduate level course in Personality Development and Mental Hygiene at the Uni-

versity of Minnesota during the summer of 1960. Seventy-six were males and 39 were females. Most subjects had had professional experience in one or more phases of education; 5 were public health nurses and 10 had had no professional experience. The average number of years of experience was 6.6 years and the range was from 0 to 30 years.

Subjects were assigned alphabetically to one or another of three groups (1, 2, or 3). The group number signified the reading set to be followed during a period of one week. This method of assignment to groups resulted in no biases due to sex, professional field, or number of years of experience. Some members of the original sample canceled the course or otherwise failed to participate in certain phases of the study, making the size of the three groups slightly unequal.

Description of the Reading Sets

The instructions of the memory, evaluation, and creative application sets were given to the subjects as follows:

Set A—Memory. When you read, it is important that you read with a set or intent to remember everything you read. In reading your assignments for the next three days, practice using this set. Try to make your mind act as a "sponge" and "soak up" as much as possible of whatever you read. Use whatever devices you like for trying to remember what you read. It may take some practice before you are really successful in assuming this set, but do not be discouraged. By the third day, you should find it easy to assume this set.

Set B—Evaluation. When you read, it is important that you think about what you are reading and evaluate critically the content. In reading your assignments for the next three days, practice using this set. As you read, ask: "Is what the author is saying really true? Does what he is saying agree with what I have experienced? Does what he is saying agree with what is known through scientific research?" Try to find defects in the reasoning of the author, in his factual presentation, and the like. Do not be concerned about errors in grammar, punctuation, sentence structure, and the like. It may take some practice ... (etc., same as above).

Set C—Creative Application. When you read, it is important that you think about the many uses of the information which you are reading. It is especially important that you think of the various ways in which the information could be used in your personal and professional life. In reading, do not just ask, "What is the author saying?" Also ask, "How can I use what the author is saying?" Do not stop with just one use. Think of as many uses as you can of the important ideas presented. Jot down some of these uses for future reference. It may take some practice ... (etc.).

Experimental Procedure

At the beginning of the first experimental week, subjects received reading-set instructions in accord with the group to which they had been assigned. Each subject received only one set during any one week. Thus, during the first week subjects in Group 1 received Set A; those in Group 2, Set B; those in Group 3, Set C; and so on —sets being shifted from group to group over the 3 weeks.

Each week each subject was provided an Evaluation of Reading Set Form upon which he was asked to report the degree to which he was able to maintain the assigned set, the degree of difficulty experienced in using that set, the number of minutes spent in completing the reading assignment for the last day of the week, and semantic reactions to the experience. Osgood's (1952) semantic differential technique was used in obtaining the latter. The following polar-adjective pairs were used with seven-point scales: good-bad, pleasurable-painful, labored-easy, wise-foolish, active-passive, fast-slow, cautious-rash, new-old, and interesting-boring.

Each Friday subjects were administered a 20-minute test which covered the reading assignment for that day. This test included items which were defined as: cognitive (multiple-choice), memory (fill-in-the-blanks), divergent thinking (extensions from given facts, uses of facts, questions, hypotheses, etc.), and evaluation (decision making situations and justifications of decisions). It was at this time that the Evaluation of Reading Set Forms were collected.

Part-scores based on performance on these subtests were determined for all subjects and were reported to each subject at the beginning of the next week.

The following are samples of items presumed to involve cognition:

According to a recent study by two New York scientists involving tests of physical fitness, which of the following findings was obtained?

——American children superior to European children.

——European children superior to American children.
——American children superior to Russian children.
——Russian children superior to American children.
——American children superior to Australian children.

Which of the following is *least likely* to contribute to the prevention of paranoid behavior?

——Training in straight thinking.
——Fostering attitudes of good sportsmanship.
——Finding available objects to blame.
——Developing self-ennobling attitudes toward sex.
——Facilitating social interaction.

The following are examples of the memory items:

According to Wang, the primary factor in personality maladjustment is either ————or————of the child's need for parental affection and sympathetic handling.

The experimenter who used a mechanical mother in laboratory studies at the University of Wisconsin is————.

The following illustrate the creative-application problems:

If hunger needs are not satisfied during childhood, how is this likely to affect personality development? Assume that the child has been given enough food for subsistence but rarely enough to satisfy hunger. What kind of adult personality would you expect?

Research has shown that lower socioeconomic individuals have limited ability to delay gratification. List as many ways as you can think of for using this information in educational situations to promote healthy personality adjustment.

The following is a sample of the evaluative problems:

Assume that you are a member of the controlling board of the Minneapolis General Hospital which is considering the establishment of a children's psychiatric unit for mentally disturbed children. The treatment program will be quite similar to that described in the text for the Children's Psychiatric Unit at the University of Michigan and it is estimated that the cost of treatment per child will be $6,500. In deciding this issue, how would you vote?

——Yes ——No

How would you justify your vote? Cite as many sound reasons as you can.

Analysis of Data

Analyses of variance were computed, comparing the effects of the reading sets of each of the different kinds of test performance (cognitive, memory, divergent thinking, and evaluation) for each of the 3 weeks. The evaluative-type item, unfortunately, was used in only one of the tests. Analyses of variance were computed for success in maintaining set, difficulty in maintaining set, number of minutes spent in completing the reading assignment, and each of the nine semantic differential scales for the three sets. Analyses of variance were also computed on Test Scores × Weeks and for Weeks × Groups in order to determine relative performance by weeks and by groups.

Results

The data comparing scores on the memory-type items are presented in Table 1. It will be noted that only during the third week were the differential effects of the three sets statistically significant at the .05 level of significance or better, the memory set producing the highest score. The predicted trend occurred during the second week but was not statistically significant.

Table 2 presents the means and analysis of variance data for scores on the Creative-Applications or divergent-thinking items. It will be noted that the differential effects are rather strong for each of the 3 weeks and that in all cases those working under the creative set achieved the highest mean score. Those using the evaluative set consistently achieved the next highest mean score. On the third test all three groups tended to achieve higher scores than during the earlier weeks. This is probably due to the accumulative effects of the course which emphasized creative thinking concerning the course content.

The data concerning performance on the evaluative (decision making) items for the second week are shown in Table 3. The differential effects are statistically significant, with the eval-

TABLE 1

COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE IN MEMORY TYPE TESTS (COMPLETION) FOR EACH OF THREE WEEKS

Week	Reading set	*N*	*M*	Between *SS*	Within *SS*	*F* ratio
First	Memory	40	6.04			
	Evaluative	38	6.24	3.9259	195.7898	2.226
	Creative	36	5.78			
Second	Memory	33	7.32			
	Evaluative	41	6.65	9.8958	420.2684	2.543
	Creative	37	7.20			
Third	Memory	38	8.30			
	Evaluative	34	8.15	12.7482	405.2544	3.429*
	Creative	40	7.54			

* $p < .05$.

TABLE 2

COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE IN CREATIVE-APPLICATION TYPES OF TESTS FOR EACH OF THREE WEEKS

Week	Reading set	*N*	*M*	Between *SS*	Within *SS*	*F* ratio
First	Memory	40	4.25			
	Evaluative	38	5.43	72.4602	643.6542	12.496**
	Creative	36	6.18			
Second	Memory	33	4.45			
	Evaluative	41	4.94	262.6431	372.4977	76.1493**
	Creative	37	7.96			
Third	Memory	38	6.88			
	Evaluative	34	8.37	66.2446	524.4794	13.7613**
	Creative	40	8.60			

** $p < .01$.

TABLE 3

COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE IN EVALUATIVE (DECISION MAKING) TYPE OF TEST FOR SECOND WEEK

Reading set	*N*	*M*	Between *SS*	Within *SS*	*F* ratio
Memory	33	5.88			
Evaluative	41	7.34	39.1292	363.4517	11.469**
Creative	37	6.70			

** $p < .01$.

uative, creative, and memory sets ranking in that order. Although this type of item was used at other times during the course, the only test in the experiment containing such an item was the one for the second week.

Table 4 presents the data for the cognitive items for each of the three tests. It had been anticipated that subjects using the memory set would perform best on these items. This occurred only during the second week. The differential effects were also significant for the last test, but subjects working under the evaluative set achieved the highest mean score. It will also be noted that subjects using the evaluative set also had a slight edge during the first week but the differential effects were not statistically significant.

TABLE 4

COMPARISON OF EFFECTS OF THREE READING SETS ON PERFORMANCE ON COGNITIVE TYPE OF TESTS (MULTIPLE-CHOICE) FOR EACH OF THREE WEEKS

Week	Reading set	*N*	*M*	Between *SS*	Within *SS*	*F* ratio
First	Memory	40	6.21			
	Evaluative	38	6.52	1.9062	236.9290	0.893
	Creative	36	6.33			
Second	Memory	33	4.03			
	Evaluative	41	3.32	10.3007	200.9297	5.537**
	Creative	37	3.43			
Third	Memory	38	4.68			
	Evaluative	34	5.06	7.3804	211.8738	3.797*
	Creative	40	4.42			

* $p < .05$.
** $p < .01$.

Test Performance by Group and by Week

Analyses of variance were also performed to study differential test performance by group and by week. It was found that the effects for groups are not statistically significant. The effects for weeks are statistically significant, total mean scores being higher for the third week than for the first two weeks. Thus, there is no evidence of the superiority of any one sequence of sets over the others.

Other Reactions to Assigned Sets

Failure to maintain an assigned set is obviously an important consideration in determining the effectiveness of a set in producing the desired results. Two questions on the Evaluation of Reading Set Form were designed to supply information concerning this problem. Responses concerning the completeness with which the set was maintained ranged from "not at all" (Point 1) to "almost all of the time" (Point 7). Responses concerning difficulty in learning to use the set ranged from "impossible" (Point 1) to "extremely easy" (Point 6). Weights were assigned arbitrarily. Means and *F* ratios are presented in Table 5 for the variables: success in using set, difficulty in using set, number of minutes spent in reading assignment for day, and for each of the semantic scales.

From Table 5 it will be noted that the sets had significant differential effects on both success and difficulty in using set but not on time spent on reading assignment. The memory set seems to have been the most successfully and easily maintained and the creative set seems to have been the least successfully and easily maintained. Subjects estimated that they maintained the memory set on the average about two-thirds of the time, finding it between "fairly difficult" and "fairly easy" to maintain. They reported that they were successful in using the creative set less than half of the time.

From Table 5 it will also be noted that on only four of the nine polar-adjective pairs are reactions significantly different among the three conditions of reading set. Those pairs are: fast-slow, cautious-rash, new-old, and interesting-boring. Subjects using the memory set tended towards fast, cautious, old, and boring. Those using the

TABLE 5

COMPARISON OF EFFECTS OF THREE READING SETS ON SUCCESS IN USING SET, DIFFICULTY IN USING SET, TIME SPENT ON READING ASSIGNMENT, AND SEMANTIC REACTION FOR THREE WEEKS COMBINED

Variable	Memory		Evaluative		Creative		*F* ratio
	N	*M*	*N*	*M*	*N*	*M*	
Success in using set	106	5.18	109	4.06	105	3.79	63.35**
Difficulty in keeping assigned set	106	4.54	110	4.10	105	3.97	26.72**
Number of minutes spent on reading assignment	104	112.27	110	114.52	105	112.00	0.23
Good-Bad	105	2.47	109	2.92	102	2.59	0.59
Pleasurable-Painful	103	3.08	107	3.16	103	3.01	0.87
Labored-Easy	103	3.96	107	3.64	103	3.73	2.44
Wise-Foolish	103	2.45	109	2.37	102	2.36	0.30
Active-Passive	103	2.64	107	2.67	102	2.48	1.46
Fast-Slow	102	3.96	107	4.38	103	4.33	5.02**
Cautious-Rash	102	2.86	107	2.97	102	3.13	3.27*
New-Old	103	4.55	107	3.64	103	2.88	18.38**
Interesting-Boring	103	2.76	107	2.25	102	2.13	17.11**

* $p < .05$.
** $p < .01$.

evaluative set tended towards slow and held to a middle position on the other pairs. Under the creative set the direction was towards rash, new, and interesting.

In order to determine whether or not there are differential effects with the passage of time, a comparison was made by weeks on the variables listed in Table 5. Only four of the variables gave rise to changes with time as data of analyses of variance showed. During the second week subjects tended to report greater success in maintaining the sets than during the other weeks. They also tended to spend somewhat more time on the reading assignment during this particular week. Finally, there was a consistent change towards fast and boring over the 3 weeks. Of these four variables the time spent on the reading assignment was significant at $.01 < F < .05$. The other three were significant at $<.01$ level.

DISCUSSION

From the data which have been presented it seems clear that instructing mature students to assume various reading sets will have differential effects on the kind of goals achieved. This would seem to be especially true if creative thinking and evaluative thinking are desired as outcomes. Although the students in the present study were able to maintain the creative and evaluative sets only about half of the time, attempts to maintain these sets had measurable effects on performance on tests covering the reading assignment.

Since most examinations given to large classes are of the cognitive (multiple-choice) type, it is worthwhile to note that the memory set does not uniformly result in the best performance on this type of test. In fact, in only one of the 3 weeks does the memory set result in superiority on the cognitive type of test. The evaluative set

tended to result in better mean scores during the other 2 weeks.

Because of the differential effects of the three sets on the four types of tests, one might also infer that instructors need to develop tests designed to measure different kinds of achievement, depending upon course objectives. If instructors expect their students to think imaginatively and evaluatively about what they learn, it is extremely doubtful that such goals are adequately assessed by the traditional multiple-choice and completion items.

The data indicate that students are more successful in maintaining the memory set and find it easier to maintain this set than the creative and evaluative sets. This probably results from the fact that they have had much more practice with this set than with the others. This hypothesis receives some support from reactions to the semantic differential. For example, subjects under the memory set tend to rate the experience in the direction of cautious, old, and boring. It is likely that, with practice, students will find it easier to maintain creative and evaluative sets and that better results will be achieved than those reported herein.

Finally, in interpreting the results, it should be recognized that the subjects are graduate students and that almost all of them have had rather extensive experience in various types of educational work.

Summary

This experiment was designed to test the differential effects of memory, evaluative, and creative "sets" in reading assignments in a graduate level course on Personality Development and Mental Hygiene. The 115 subjects were arranged alphabetically and assigned to one of three groups. Each group was assigned in turn each of the three reading sets to be used in the assigned readings for a week. At the end of each week, a 20-minute test containing cognitive, memory, creative thinking, and evaluative items was administered. Subjects also estimated their degree of success in maintaining the set, difficulty in maintaining it, the number of minutes spent on the Friday's reading assignment, and their semantic reactions to nine polar-adjective pairs.

Analyses of variance indicated that the three sets produced differential effects on most of the tests for all 3 weeks. Those operating under the creative set, achieved the highest mean each week on the creative applications. Those operating under the evaluative set achieved the highest mean on the evaluative or decision making problem when such a problem was given (second week only). Those operating under the memory set achieved the highest mean on the memory tests (completion) only during the third week. The differential effects on the memory items were not statistically significant during the other weeks. Differential effects on the cognitive items were not consistent. Those operating under the memory set achieved the highest mean during the second week and those under the evaluative set led during the third week.

Subjects reported that they found it easier to maintain the memory set and rated themselves as more successful in maintaining this set than the other two. There were no differential effects on the estimated length of time spent on the reading assignment. On the semantic differential, subjects under the memory set tended towards fast, cautious, old, and boring. Those

under the evaluative set tended towards slow and those under the creative set, towards rash, new, and interesting.

During the second week subjects generally reported greater success in maintaining sets than during the first and third weeks. On the semantic evaluations, there was a consistent and significant drift towards fast and boring. The three original groups did not differ significantly, giving no evidence of the superiority of any one sequence of sets over the others.

REFERENCES

GUILFORD, J. P. Three faces of intellect. *Amer. Psychologist*, 1959, **14**, 469–479.

JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harpers, 1955.

OSGOOD, C. E. The nature and measurement of meaning. *Psychol. Bull.*, 1952, **49**, 192–237.

TORRANCE, E. P. The phenomenon of resistance in learning. *J. abnorm. soc. Psychol.*, 1950, **45**, 592–597.

TORRANCE, E. P. Getting mental-hygiene practices into action through a college class. *Ment. Hyg., NY*, 1951, **35**, 88–95.

TORRANCE, E. P. Effects of induced evaluative sets on the development of new ideas among graduate students. In E. P. Torrance (Ed.), *Creativity: Second conference on gifted children*. Minneapolis: University of Minnesota, Center for Continuation Study, 1960.

(Received February 8, 1961)

Journal of Educational Psychology
1961, Vol. 52, No. 4, 215-221

THE EFFECTS OF VISUAL DISCRIMINATION PRETRAINING WITH WORD AND LETTER STIMULI ON LEARNING TO READ A WORD LIST IN KINDERGARTEN CHILDREN

SIEGMAR MUEHL[1]

Iowa Child Welfare Research Station

Ability to discriminate among word forms is a prerequisite to associating these same stimuli with appropriate word names.

An earlier study by the author (Muehl, 1960) using kindergarten subjects, showed that visual discrimination pretraining with the same words to be read in a subsequent word list facilitated learning the word list when compared to discrimination pretraining with different words or pretraining with geometric forms. These results suggested the possibility that specific observing responses were learned to the stimuli in the same-word pretraining; and, that these responses transferred positively to discriminating among the words in the vocabulary task.

The purpose of the present experimental study was to determine, more specifically, the stimuli relevant to visual discrimination among word forms for beginning readers.

One assumption tested is that the child learns to discriminate among words on the basis of the shape or form of the whole word. A second assumption tested is that the child discriminates among words on the basis of attending to some part of the word, such as a particular letter, part of a letter, or letter grouping.

The assumption that word shape is the basis for visual discrimination among words, although lacking objective confirmation (Davidson, 1931), appears to be supported by educators, judging by the widespread practice of the so-called "whole-word" method of introducing reading in the first grade. According to Gray (1956, pp. 83–84), this method is based on the assumption that, "... individual words are the basic units of thought and recognition," or, that, "... each word has a characteristic form by which it can be remembered." Although words are the basic meaning units in our language, it does not follow they are also the basic units of visual recognition.

The assumption that the beginning reader discriminates among words on the basis of letters finds support in two different areas of educational research. Comparing intralist errors in word recognition tasks for children ranging from 4 to 6 years in age, Gates and Boeker (1923), Meek (1925), and Muehl (1960) found that letter similarities were the apparent source of confusion among words. Using a camera to record eye movements, Buswell (1922) showed that beginning readers made many eye fixations associated with details within a word in the early reading stages.

The results of psychological re-

[1] The author wishes to thank Buford W. Garner, Superintendent of Schools, Iowa City; Marjorie Clatterbaugh, Donald Tvedt, and John Moore, principals; and Nancy Cooker, Audrey Nelson, Nellie Reger, and Carol Ripp, kindergarten teachers whose cooperation made this research possible.

search bearing on the related question of part vs. whole discrimination are inconclusive. A review of this research with children, reported by Gibson and Olum (1960), showed that the research results were conflicting and appeared to depend on the stimulus material used. However, none of the studies reviewed had used word stimuli.

Method

Pretraining Task

To test these alternative assumptions, three groups of kindergarten children were differentiated by the kinds of visual discrimination or matching pretraining each received prior to learning to read a word list. The pretraining paradigm is shown in Table 1. The nonsense word sets for Groups RR and IR were designed so that either word shape (stimulus configuration determined by all the letters) or letter differences among the three words in each set could serve as relevant stimuli for discriminating among the words. LD pretraining was included to control for the effect of learning to discriminate among the relevant letters, as such, on later word list performance.

Each of the pretraining sets (RR, IR, and LD) was presented in four orders. Two of these orders presented consecutively constituted a two-trial block. The words and letters appearing in the sets were typed in lower case with a primary-style typewriter. The word stimuli were approximately 11 mm. in length and 5 mm. in height to include the tallest letters. These word and letter stimuli were mounted in the center of 2½″ × 2¾″ white cards. Accompanying each two-trial block of stimulus items in the three sets were two, 2¾″ × 8½″, multiple-choice response cards. The three words, or three letters, appearing on the stimulus cards in each set also appeared on these response cards, in two different position orders. The apparatus for presenting the stimulus and response cards has been described in detail elsewhere (Muehl, 1960). Briefly, it consisted of a tray for presenting the stimulus cards, a hinged window below the tray for exposing the response card, and three buttons located on a panel immediately in front of each of the three words appearing on the exposed response card. If the subject selected the correct response word and pressed the button under it a door chime would ring. The experimenter controlled the appropriate button-chime connection by means of a selector switch from behind the apparatus.

All subjects received eight matching trials. A presentation consisted of the experimenter's placing a stimulus card on the

TABLE 1

Stimulus-Response Paradigm to Illustrate Experimental Design

Visual Discrimination Pretraining		Reading Task	
Stimuli	Responses	Stimuli	Responses
Group RR[a]		All Groups	
feu	All Groups	feu	blue
geu		geu	red
reu	matching	reu	white
Group IR[b]			
fjd			
gjd			
rjd			
Group LD[c]			
f		LD	same as above
g			
r		LD-O	same as above, plus orienting instructions

[a] Relevant shape-relevant letter.
[b] Irrelevant shape-relevant letter.
[c] Letter discrimination.

card-holder tray for approximately 2 seconds. The subject was told to look carefully at the card. The experimenter then removed the card and the subject was instructed to open the response card window, find the same stimulus item on the response card that had appeared on the stimulus card, and then press the button directly under it to find if his choice was correct. If the subject pushed the incorrect button on any one of the first three presentations, the experimenter showed the stimulus item again and had the subject find the correct response item and ring the bell. On subsequent presentations, if the subject responded incorrectly or if the subject did not respond in approximately 5 seconds, the experimenter closed the response card window and presented the next stimulus card. The experimenter alternated response cards with each two-trial block. No rewards were given in pretraining other than provided by the chime, and verbal encouragement from the experimenter.

Reading Task

The reading or transfer task was designed to make both word-shape and letter differences among the Group RR pretraining words relevant stimuli for discriminating among the reading task words. For Group IR pretraining words, the reading task was designed to make letter differences alone relevant to discriminating among the reading task words. This reading task paradigm is shown in Table 1.

Set RR nonsense words were paired with ⅝″ × ⅞″ color patches. The typed words and color patches were mounted on 3½″ × 6¼″ white, plastic cards for paired-associate presentation.

The apparatus used for the reading task was a Hunter Card Master. The Card Master is designed to present automatically the above plastic cards before a 3″ × 5½″ aperture. By means of a timer unit and two metal shutters, the aperture opening can be controlled to secure the following exposure intervals: an anticipation interval to expose the stimulus word alone; a joint presentation interval to expose the word-color pair; and a between-item interval when the aperture is completely covered, during which time the next stimulus-response pair is brought into position behind the shutters.

Four orders of the three word-color pairs appeared in the reading task. Two of these orders presented consecutively constituted a two-trial block. A blank card appeared at the end of each two-trial block. The following presentation intervals were used: anticipation interval—3 seconds; joint presentation interval—3 seconds; and a between-item interval—2 seconds. All subjects continued in the reading task to a learning criterion of 11 of 12 correct anticipations in two successive two-trial blocks, or for a maximum of 20 trials. Subjects reaching a learning criterion before 20 trials were given perfect scores on the remaining trials.

The reading task was administered immediately following pretraining. At the beginning of the reading task, the subject was shown each color patch with the word concealed and told the appropriate color name. The three color patches were then presented a second time and the subject asked to name the colors. If the subject could not give the correct color names he was eliminated. Following the color naming, a familiarization trial with the three word-color pairs was given. The subject was told with each paired presentation that the word and color went together. After presenting the third pair the subject was instructed that the game was to guess the color that went with each word before the color appeared. The learning trials began following these instructions for subjects in Groups RR, IR, and for half the LD group. Since the LD group would not see the relevant letters as parts of words during pretraining, subjects in this group, presumably, would need to learn where to look for these letters in the reading task words, as well as learn to discriminate the relevant letter from the rest of the word. To control for learning where to look, the other half of the LD group, designated LD-O, was given special orienting instructions at the beginning of the reading task. At the end of the familiarization trial, the experimenter told subjects in the LD-O group always to look at the beginning of the word and this would help them learn the right color. At the same time the experimenter pointed to the beginning of the word. These instructions were repeated twice more, at the end of trial Blocks 1–2 and 3–4.

To secure attention to the stimulus words as they appeared in the aperture during the reading task, the experimenter frequently pointed to the window with a pencil on each presentation. If the subject did not respond during the anticipation interval, the experimenter encouraged the subject to try to guess a color after looking carefuly at each word.

If a correct response occurred in the an-

ticipation interval, the subject moved a bead across a counting frame. If an incorrect response or omission occurred, the experimenter named the correct color in the joint presentation interval. The bead incentive and occasional verbal encouragement from the experimenter served as rewards in the reading task.

Subjects

The subjects were 60 children from three kindergartens in the public schools in Iowa City. The subjects were randomly assigned to the three pretraining groups: 15 each to Groups RR and IR, 30 to Group LD. Three additional subjects were eliminated from the study because they could not respond correctly in naming the color patches. Subjects ranged in age from 61 to 72 months. The mean age was 66.7 months. There were no significant age differences among the three pretraining groups, either in variability, or mean age. Testing was done in the public schools during November and December.

TABLE 2

MEAN NUMBER OF CORRECT RESPONSES DURING VISUAL DISCRIMINATION PRETRAINING

Group	Blocks of two trials				
	1-2	3-4	5-6	7-8	Total
WD[a]	4.27	4.60	5.03	5.10	19.00
LD	5.47	5.23	5.43	5.57	21.70

[a] Groups RR and IR combined.

TABLE 3

ANALYSIS OF VARIANCE OF CORRECT RESPONSES IN EIGHT LEARNING TRIALS FOR GROUPS WD AND LD IN VISUAL DISCRIMINATION PRETRAINING

(Two-trial blocks)

Source	*df*	*MS*	*F*	*p*
Between-Subjects	59			
Groups (B)	1	27.34	11.02	<.005
Error (b)	58	2.48		
Within-Subjects	180			
Trials (A)	3	3.19	4.25	<.01
A × B	3	1.98	2.64	<.10
Error (w)	174	.75		
Total	239			

RESULTS

Pretraining

Comparing Groups RR and IR, the mean number of correct responses were 19.40 and 18.60, respectively; the standard deviations 4.22 and 3.52, respectively. A Lindquist (1956) Type I analysis of variance of the pretraining scores showed no reliable performance differences between these groups. The trials effect for the combined groups was significant ($p < .01$), indicating a learning effect. Since there were no performance differences between Groups RR and IR in pretraining these two word discrimination groups were combined and designated as Group WD. Comparing Groups WD and LD in pretraining, the mean number of correct responses were 19.00 and 21.70, respectively; the standard deviations 3.91 and 1.98, respectively. Hartley's test (Walker & Lev, 1953, pp. 193–194) for homogeneity of variance between the groups was significant ($p < .01$), leading to the rejection of the null hypothesis. The mean numbers of correct responses for Groups WD and LD during pretraining are presented in Table 2. A summary of a Lindquist (1956) Type I analysis of variance for these data is presented in Table 3. As shown in the summary, differences in group means and the trials effect were significant ($p = .025$, to adjust for effects of heterogeneity of variance). A further analysis of simple effects showed that the trials or learning effect was entirely accounted for within the WD group ($p < .01$).

Reading Task

Comparing Groups RR and IR, the mean number of correct responses were 33.07 and 32.93, respectively; the

TABLE 4

MEAN NUMBER OF CORRECT RESPONSES AND STANDARD DEVIATIONS FOR GROUPS ON THE READING TASK

Group	Mean	*SD*
WD[a]	33.00	11.67
LD	29.53	12.75
LD-O	32.53	14.59

[a] Groups RR and IR combined.

learning curves in the reading task for the three groups are presented in Figure 1. A summary of a Lindquist (1956) Type I analysis of variance for these data showed no significant overall performance difference for the groups on the reading task. Comparing Group LD with Groups WD and LD-O separately showed no reliable performance differences ($F < 1$). Performance differences on the first trial block also were not significant.

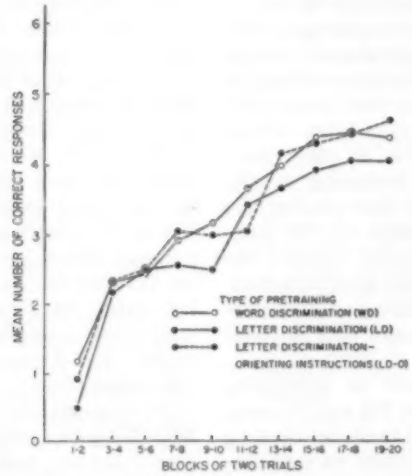

FIG. 1. The effects of visual discrimination pretraining with relevant letters contained in words (Group WD: Groups RR and IR combined) vs. relevant letters presented alone (Groups LD and LD-O) on correct responses in learning to read a three word list.

standard deviations, 11.58 and 11.75, respectively. A Lindquist (1956) Type I analysis of variance showed Groups RR and IR did not differ significantly in overall performance on the reading task. A further comparison between the groups on the first trial block also showed no reliable differences.

On the basis of these findings, the two word-discrimination pretraining groups were again combined and designated as Group WD in the reading task. The mean number of correct responses for Groups WD, LD, and LD-O are presented in Table 4. The

DISCUSSION

If word shape is the relevant visual aspect for discriminating among words for beginning readers, Group RR pretraining should have resulted in transfer of discrimination learning to the reading task and faster learning of the list relative to Group IR pretraining. On the other hand, if the relevant letters alone provide the basis for visual discrimination in pretraining, Groups RR and IR pretraining should

have resulted in equal transfer of discrimination learning and no performance difference between the groups on the reading task.

The findings based on comparing Groups RR and IR on the reading task are consistent with the assumption that beginning readers discriminate among similar length words of different shape on the basis of specific letter differences. Further investigation is necessary to determine the effects of at least two other word-stimulus variables on visual discrimination learning: word-shape differences defined in terms of differences in word length and location of the relevant letter within the stimulus word, i.e., in other positions than at the beginning of the word. It is possible that placing the relevant letter in some middle position in the word would increase the difficulty of the letter discrimination. This effect could increase the importance of shape or length cues as bases for discrimination.

The findings based on comparing Groups WD (Groups RR and IR combined), LD, and LD-O on reading task performance require an explanation to account for the appearance of learning differences between the word and letter discrimination groups in pretraining, and the absence of learning differences on the reading task. Learning differences in discrimination pretraining indicated differential difficulty in discriminating letters when these letters were presented singly as compared to being presented in words. Absence of learning differences on the reading task indicated that the addition of irrelevant letters to the previously differentiated single-letter stimuli for Groups LD and LD-O did not increase the difficulty of discriminating these letters.

To account for the relative difficulty of the WD group discrimination learning in pretraining, it is assumed that generalization among the nonsense words increased as a result of the addition of similar stimulus elements—the same two letters—to the relevant letters in each of the words. Thus to discriminate among the words, WD group subjects had to acquire a set of observing responses related either to the location of the relevant letter in the words, discriminating the relevant letter from the total stimulus complex, or a combination of the two types of observing responses.

On the basis of this reasoning, however, differences in learning performance should have appeared in the reading task, at least for the LD group. The assumption would be that the LD group would need to acquire the same orienting responses to the relevant letters included in the reading task words. This additional learning would be predicted to interfere with the acquisition of the word names. The findings did not confirm this expectation.

To account for the absence of a WD-LD group performance difference on the reading task, two explanations are offered. The simplest explanation is based on the greater difficulty and variability of the reading task. The increase in performance variability could mask any differential transfer of discrimination learning between the pretraining groups. A second explanation requires the assumption that exposure to the relevant letters in the LD group pretraining resulted in an "acquired distinctiveness" of the letter stimuli, not detected by the pretraining response measure, that transferred to the reading task. Transfer of this acquired distinctiveness could facilitate the acquisition of the observing responses necessary to discriminate the relevant letters included in the reading task words—the same

observing responses acquired more slowly by the WD group in pretraining.

The second explanation suggests the hypothesis that although letters included in words are more difficult to discriminate than the same letters presented singly, this relative difficulty in learning to discriminate letters in words can be overcome by providing visual discrimination training with the relevant letters prior to presenting them as parts of words.

Further research is necessary to test this hypothesis when the relevant letters are included in other than the first position in the word.

Summary

The purpose of this study was to determine the stimuli relevant to visual discrimination among word forms for beginning readers.

Of three groups of kindergarten children, two received discrimination pretraining with words. For one group, the pretraining words were constructed so that both shape and letter differences among the words were relevant to discriminating among the words in the reading task. For the second group, only the letter differences among the pretraining words were relevant to discriminating among the reading task words. A third group received discrimination pretraining with the relevant letters alone.

There were no significant performance differences between the two word discrimination groups on either a pretraining or reading task, a finding consistent with the assumption that children discriminate among similar length words of different shape on the basis of specific letter differences.

When the letter discrimination pretraining group was compared with the combined word discrimination pretraining groups, there was a performance difference favoring the letter discrimination group in pretraining, but no differences on the reading task between the groups. These findings suggested the hypothesis that although letters included in words are more difficult to discriminate than the same letters presented singly, this relative difficulty in learning to discriminate letters in words can be overcome by providing visual discrimination training with the relevant letters prior to presenting them as parts of words.

REFERENCES

Buswell, G. T. Fundamental reading habits: A study of their development. *Suppl. educ. Monogr.*, 1922, No. 21.

Davidson, Helen P. An experimental study of bright, average, and dull children at the four-year mental level. *Genet. psychol. Monogr.*, 1931, **9**, 119–289.

Gates, A. I., & Boeker, Eloise. A study of initial states in reading by preschool children. *Teach. Coll. Rec.*, 1923, **24**, 469–488.

Gibson, Eleanor J., & Olum, Vivian. Experimental methods of studying perception in children. In P. H. Mussen (Ed.), *Handbook of research methods in child psychology*. New York: Wiley, 1960. Pp. 311–373.

Gray, W. S. *The teaching of reading and writing: An international survey*. (Monographs on Fundamental Education, No. 10) New York: UNESCO, 1956.

Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1956.

Meek, Lois H. A study of learning and retention in young children. *Teach. Coll. Contrib. Educ.*, 1925, No. 164.

Muehl, S. The effects of visual discrimination pretraining on learning to read a vocabulary list in kindergarten children. *J. educ. Psychol.*, 1960, **51**, 217–221.

Walker, Helen M., & Lev, J. *Statistical inference*. New York: Holt, 1953.

(Received February 9, 1961)

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.